

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: 2003年11月26日

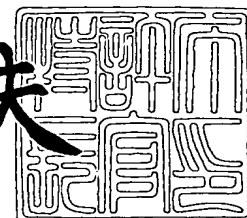
出願番号
Application Number: 特願2003-394922
[ST. 10/C]: [JP 2003-394922]

出願人
Applicant(s): 株式会社日立製作所

2004年 3月 2日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



出証番号 出証特2004-3015586

【書類名】 特許願
【整理番号】 K03017521A
【あて先】 特許庁長官殿
【国際特許分類】 G06F 3/06
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A
 I D システム事業部内
 【氏名】 小笠原 裕
【発明者】
 【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所
 ソフトウェア事業部内
 【氏名】 蟹江 誉
【発明者】
 【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所
 ソフトウェア事業部内
 【氏名】 雑賀 信之
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A
 I D システム事業部内
 【氏名】 ▲高▼田 豊
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A
 I D システム事業部内
 【氏名】 中山 信一
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社 日立製作所
【代理人】
 【識別番号】 100075096
 【弁理士】
 【氏名又は名称】 作田 康夫
【選任した代理人】
 【識別番号】 100100310
 【弁理士】
 【氏名又は名称】 井上 学
【手数料の表示】
 【予納台帳番号】 013088
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1

【書類名】 特許請求の範囲**【請求項 1】**

データを格納する複数の記憶デバイスと、
前記複数の記憶デバイスに対するデータの格納を制御する記憶デバイス制御部と、
前記記憶デバイス制御部に接続される接続部と、
自ディスクアレイ装置の外部のローカルエリアネットワークを介して受けたファイルレベルのデータをブロックレベルのデータに変換して、前記複数の記憶デバイスへの格納を要求する第一のプロセッサと、前記第一のプロセッサからの要求に応じて前記接続部及び前記記憶デバイス制御部を介して前記複数の記憶デバイスへ前記ブロックレベルのデータを転送する第二のプロセッサとを有し、前記接続部及び前記ローカルエリアネットワークに接続される複数の第一のチャンネル制御部と、
前記複数の第一のチャンネル制御部及び前記記憶デバイス制御部によってやり取りされる制御情報が格納される共有メモリと、
前記複数の第一のチャンネル制御部と前記記憶デバイス制御部との間でやり取りされるデータを一時的に保存するキャッシュメモリと、を有し、
前記複数の第一のチャンネル制御部内の前記第二のプロセッサは、前記ブロックレベルのデータが格納される複数の記憶領域と、複数の前記第一のプロセッサによって相互にやり取りされるプロセッサ間の処理状況に関する情報が格納されるプロセッサ情報格納領域と、を前記複数の記憶デバイスの記憶領域を用いて作成するものであり、
前記記憶デバイス制御部は、前記複数の第一のチャンネル制御部内の前記第一のプロセッサの指示に応じて、前記プロセッサ情報格納領域に格納された情報を、前記複数の記憶デバイスの記憶領域を用いて作成されたプロセッサ情報バックアップ用の格納領域に対してコピーするように制御するものであることを特徴とするディスクアレイ装置。

【請求項 2】

請求項 1 に記載のディスクアレイ装置において、
前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記第一のプロセッサが設けられている前記第一のチャンネル制御部内の前記第二のプロセッサに対して、前記第一のプロセッサの処理状況に関する情報を前記プロセッサ情報格納領域に格納するように指示するものであり、
前記第一のプロセッサが設けられている前記第一のチャンネル制御部内の前記第二のプロセッサは、前記第一のチャンネル制御部の指示に応じて、前記第一のプロセッサの処理状況に関する情報を、前記プロセッサ情報格納領域に格納するように制御するものであることを特徴とするディスクアレイ装置。

【請求項 3】

請求項 1 に記載のディスクアレイ装置において、
前記複数の第一のチャンネル制御部内の第二のプロセッサは、前記第二のプロセッサが設けられている前記第一のチャンネル制御部内の前記第一のプロセッサの要求に従って、前記ブロックレベルのデータを前記キャッシュメモリに保存するとともに、前記ブロックレベルのデータを前記キャッシュメモリに保存したことを表す情報を前記共有メモリに格納するものであり、
共有メモリは、前記複数の第一のチャンネル制御部内の第二のプロセッサの制御のもとに、前記ブロックレベルのデータが前記キャッシュメモリに保存されたことを表す情報が格納されるものであることを特徴とするディスクアレイ装置。

【請求項 4】

請求項 1 に記載のディスクアレイ装置において、
前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記記憶デバイス制御部に対して、前記プロセッサ情報格納領域に格納された情報を、前記プロセッサ情報バックアップ用の格納領域に対してコピーするように指示するものであり、
前記記憶デバイス制御部は、前記第一のプロセッサの指示に応じて、コピー処理を制御するものであることを特徴とするディスクアレイ装置。

【請求項 5】

請求項 4 に記載のディスクアレイ装置において、

前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記プロセッサ情報格納領域に格納された情報の読み出し又は書き込みが不可能になった場合、前記プロセッサ情報バックアップ用の格納領域に格納された情報を読み出し又は書き込むことにより、処理を継続することを特徴とするディスクアレイ装置。

【請求項 6】

請求項 1 に記載のディスクアレイ装置において、

前記複数の第一のチャンネル制御部は、複数のクラスタグループに分類されるものであり、

前記プロセッサ情報格納領域は、複数のプロセッサ情報格納部分を有するものであり、前記複数のクラスタグループは、前記複数のプロセッサ情報格納部分のうちの各々異なる部分を割当てられるものであることを特徴とするディスクアレイ装置。

【請求項 7】

請求項 6 に記載のディスクアレイ装置において、

前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部は、前記複数のプロセッサ情報格納部分のうちの第一のプロセッサ情報格納部分に対して、前記第一のプロセッサによって相互にやり取りされるプロセッサ間の処理状況に関する情報を格納するものであり、

前記複数のクラスタグループのうちの第二のクラスタグループに含まれる前記複数の第一のチャンネル制御部は、前記複数のプロセッサ情報格納部分のうちの第二のプロセッサ情報格納部分に対して、前記第一のプロセッサによって相互にやり取りされるプロセッサ間の処理状況に関する情報を格納するものであることを特徴とするディスクアレイ装置。

【請求項 8】

請求項 7 に記載のディスクアレイ装置において、

前記第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部の前記第一のプロセッサは、前記第一のプロセッサ情報格納部分に格納された情報の複製の作成を、前記記憶デバイス制御部に対して指示するものであり、

前記記憶デバイス制御部は、前記第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部の前記第一のプロセッサの指示に応じて、前記プロセッサ情報バックアップ用の格納領域に含まれる第一のバックアップ領域に対して、前記第一のプロセッサ情報格納部分に格納された情報の複製を格納するものであることを特徴とするディスクアレイ装置。

【請求項 9】

請求項 7 に記載のディスクアレイ装置において、

前記第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部の前記第一のプロセッサは、前記第一のプロセッサ情報格納部分及び前記第二のプロセッサ情報格納部分に格納された情報の複製の作成を、前記記憶デバイス制御部に対して指示するものであり、

前記記憶デバイス制御部は、前記第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部の前記第一のプロセッサの指示に応じて、前記プロセッサ情報バックアップ用の格納領域に含まれる第一のバックアップ領域及び第二のバックアップ領域に対して、前記第一のプロセッサ情報格納部分及び前記第二のプロセッサ情報格納部分に格納された情報の複製を格納するものであることを特徴とするディスクアレイ装置。

【請求項 10】

請求項 7 に記載のディスクアレイ装置において、

前記複数の第一のチャンネル制御部及び前記記憶デバイス制御部に関する情報の取得に利用される管理端末と、を有し、

前記記憶デバイス制御部は、前記管理端末からの指示に応じて、前記プロセッサ情報バックアップ用の格納領域に含まれる第一のバックアップ領域及び第二のバックアップ領域

に対して、前記第一のプロセッサ情報格納部分及び前記第二のプロセッサ情報格納部分に格納された情報の複製を格納するものであることを特徴とするディスクアレイ装置。

【請求項 1 1】

請求項 8 に記載のディスクアレイ装置において、

前記第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部の前記第一のプロセッサは、前記第一のプロセッサ情報格納部分に格納された情報の読み出し又は書き込みが不可能になった場合、前記第一のバックアップ領域に格納された情報を読み出し又は書き込むことにより、処理を継続するものであることを特徴とするディスクアレイ装置。

【請求項 1 2】

データを格納する複数の記憶デバイスと、

前記複数の記憶デバイスに対するデータの格納を制御する記憶デバイス制御部と、

前記記憶デバイス制御部に接続される接続部と、

自ディスクアレイ装置の外部のローカルエリアネットワークを介して受けたファイルレベルのデータをブロックレベルのデータに変換して、前記複数の記憶デバイスへの格納を要求する第一のプロセッサと、前記第一のプロセッサからの要求に応じて前記接続部及び前記記憶デバイス制御部を介して前記複数の記憶デバイスへ前記ブロックレベルのデータを転送する第二のプロセッサとを有し、前記接続部及び前記ローカルエリアネットワークに接続される複数の第一のチャンネル制御部と、

前記複数の第一のチャンネル制御部及び前記記憶デバイス制御部によってやり取りされる制御情報が格納される共有メモリと、

前記複数の第一のチャンネル制御部と前記記憶デバイス制御部との間でやり取りされるデータを一時的に保存するキャッシュメモリと、を有し、

前記複数の第一のチャンネル制御部内の前記第二のプロセッサは、前記ブロックレベルのデータが格納される複数の記憶領域と、複数の前記第一のプロセッサによって相互にやり取りされるプロセッサ間の処理状況に関する情報が格納されるプロセッサ情報格納領域と、複数の前記第一のプロセッサ上で動作するソフトウェアプログラムが格納されるソフトウェアプログラム格納領域と、を前記複数の記憶デバイスの記憶領域を用いて作成するものであり、

前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記第一のプロセッサが設けられている前記第一のチャンネル制御部内の前記第二のプロセッサの制御に応じて、前記ソフトウェアプログラム格納領域に格納されているソフトウェアプログラムを取得して、動作するものであることを特徴とするディスクアレイ装置。

【請求項 1 3】

請求項 1 2 に記載のディスクアレイ装置において、

前記複数の第一のチャンネル制御部内の前記第一のプロセッサで動作するソフトウェアプログラムは、前記記憶デバイス制御部に対して、前記プロセッサ情報格納領域に格納された情報を、前記プロセッサ情報バックアップ用の格納領域に対してコピーするように指示するものであり、

前記記憶デバイス制御部は、前記複数の第一のチャンネル制御部内の前記第一のプロセッサの指示に応じて、前記プロセッサ情報格納領域に格納された情報を、前記複数の記憶デバイスの記憶領域を用いて作成されたプロセッサ情報バックアップ用の格納領域に対してコピーするように制御するものであることを特徴とするディスクアレイ装置。

【請求項 1 4】

請求項 1 2 に記載のディスクアレイ装置において、

前記複数の第一のチャンネル制御部は、複数のクラスタグループに分類されるものであり、

前記プロセッサ情報格納領域は、複数のプロセッサ情報格納部分を有するものであり、

前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部で動作するソフトウェアプログラムの各々は、相互に、前記複数の

ロセッサ情報格納部分のうちの第一のプロセッサ情報格納部分に対して、プロセッサ間の処理状況に関する情報を格納しながら動作するものであることを特徴とするディスクアレイ装置。

【請求項 15】

請求項 12 に記載のディスクアレイ装置において、
前記複数の第一のチャンネル制御部は、複数のクラスタグループに分類されるものであり、

前記プロセッサ情報格納領域は、複数のプロセッサ情報格納部分を有するものであり、
前記複数のプロセッサ情報格納部分に格納された情報は、前記複数のクラスタグループ毎に、前記複数のプロセッサ情報格納部分に対応する複数のバックアップ領域に複製されるものであることを特徴とするディスクアレイ装置。

【請求項 16】

請求項 15 に記載のディスクアレイ装置において、
前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記第一のプロセッサが設けられている前記第一のチャンネル制御部内の前記第二のプロセッサを介して、前記記憶デバイス制御部に対して、ブロック単位で前記複製を実行するように指示するものであり、

前記記憶デバイス制御部は、前記第一のプロセッサの指示に応じて、前記複数のプロセッサ情報格納部分のうちの第一のプロセッサ情報格納部分に格納された情報を、ブロック単位で、前記複数のバックアップ領域のうちの第一のバックアップ領域に複製するものであることを特徴とするディスクアレイ装置。

【請求項 17】

請求項 15 に記載のディスクアレイ装置において、
前記ローカルエリアネットワークには、端末が設けられており、

前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記端末からの指示に応じて、前記第一のプロセッサが設けられている前記第一のチャンネル制御部内の前記第二のプロセッサを介して、前記記憶デバイス制御部に対して、ブロック単位で前記複製を実行するように指示するものであることを特徴とするディスクアレイ装置。

【請求項 18】

請求項 15 に記載のディスクアレイ装置において、
前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、一定の時間間隔毎に、前記第一のプロセッサが設けられている前記第一のチャンネル制御部内の前記第二のプロセッサを介して、前記記憶デバイス制御部に対して、ブロック単位で前記複製を実行するように指示するものであることを特徴とするディスクアレイ装置。

【請求項 19】

請求項 15 に記載のディスクアレイ装置において、
前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記第一のプロセッサが設けられている前記第一のチャンネル制御部内の前記第二のプロセッサを介して、前記記憶デバイス制御部の負荷状態を取得し、前記記憶デバイス制御部の負荷状態に応じて、前記記憶デバイス制御部に対して、ブロック単位で前記複製を実行するように指示するものであることを特徴とするディスクアレイ装置。

【請求項 20】

請求項 15 に記載のディスクアレイ装置において、
前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記複数のプロセッサ情報格納部分のうちの第一のプロセッサ情報格納部分に格納された情報へのアクセスが不可能になった場合に、前記第一のバックアップ領域に格納された情報を用いて処理を実行し、前記第一の

プロセッサ情報格納部分が新たに形成された場合に、前記第一のバックアップ領域に格納された情報を前記新たに形成された第一のプロセッサ情報格納部分に複製して、前記新たに形成された第一のプロセッサ情報格納部分に格納された情報を用いて、処理を実行するものであることを特徴とするディスクアレイ装置。

【書類名】明細書**【発明の名称】** ディスクアレイ装置**【技術分野】****【0001】**

本発明は、複数の異種ネットワークに接続可能なように新たに発明された記憶装置システムに関し、特に記憶装置システムの複製を制御する方法に関する。

【背景技術】**【0002】**

近年コンピュータシステムで取り扱われるデータ量が急激に増加している。かかる膨大なデータを効率よく利用し管理するために、複数のディスクアレイ装置（以下、記憶装置システムと称する）と情報処理装置とを専用のネットワーク（Storage Area Network、以下 SAN と記す）で接続し、記憶装置システムへの高速かつ大量なアクセスを実現する技術が開発されている。記憶装置システムと情報処理装置とを SAN で接続し高速なデータ転送を実現するためには、ファイバチャネルプロトコルに従った通信機器を用いてネットワークを構築するのが一般的である。

【0003】

一方、複数の記憶装置システムと情報処理装置とを TCP/IP（Transmission Control Protocol/Internet Protocol）プロトコルを用いたネットワークで相互に接続し、記憶装置システムへのファイルレベルでのアクセスを実現する、NAS（Network Attached Storage）と呼ばれるネットワークシステムが開発されている。NAS においては、記憶装置システムに対してファイルシステム機能を有する装置が接続されているため、情報処理装置からのファイルレベルでのアクセスが可能となっている。特に最近ではミッドレンジクラスやエンタープライズクラスと呼ばれるような、巨大な記憶資源を提供する RAID（Redundant Arrays of Inexpensive Disks）方式で管理された記憶装置システムにファイルシステムを結合させた、大規模な NAS が注目されている。

【0004】

【特許文献 1】 特開 2002-351703 号公報

【発明の開示】**【発明が解決しようとする課題】****【0005】**

しかしながら従来の NAS は、TCP/IP 通信機能及びファイルシステム機能を持たない記憶装置システムに、TCP/IP 通信機能及びファイルシステム機能を持った情報処理装置を接続させることにより実現されていた。そのため、上記接続される情報処理装置の設置スペースが必要であった。また上記情報処理装置と記憶装置システムとの間は、高速に通信を行う必要性から SAN で接続されていることが多く、そのための通信制御機器や通信制御機能を備える必要もあった。

【0006】

本発明は上記課題を鑑みてなされたものであり、複数の異種ネットワークに接続可能なように全く新しく発明された記憶装置システム、及びかかる記憶装置システムを発明するにあたり必要とされる記憶デバイス制御装置、及びデバイス制御装置の複製を制御する方法を提供することを主たる目的とする。

【課題を解決するための手段】**【0007】**

本発明のディスクアレイ装置は、以下の構成を有する。

【0008】

ディスクアレイ装置は、データを格納する複数の記憶デバイスと、前記複数の記憶デバイスに対するデータの格納を制御する記憶デバイス制御部と、前記記憶デバイス制御部に接続される接続部と、複数の第一のチャネル制御部と、前記複数の第一のチャネル制御部及び前記記憶デバイス制御部によってやり取りされる制御情報が格納される共有メモリと、前記複数の第一のチャネル制御部と前記記憶デバイス制御部との間でやり取りされるデ

ータを一時的に保存するキャッシュメモリと、を有する。

【0009】

第一のチャネル制御部は、自ディスクアレイ装置の外部のローカルエリアネットワークを介して受けたファイルレベルのデータをブロックレベルのデータに変換して、前記複数の記憶デバイスへの格納を要求する第一のプロセッサと、前記第一のプロセッサからの要求に応じて前記接続部及び前記記憶デバイス制御部を介して前記複数の記憶デバイスへ前記ブロックレベルのデータを転送する第二のプロセッサとを有し、前記接続部及び前記ローカルエリアネットワークに接続される。

【0010】

前記複数の第一のチャネル制御部内の前記第二のプロセッサは、前記ブロックレベルのデータが格納される複数の記憶領域と、複数の前記第一のプロセッサによって相互にやり取りされるプロセッサ間の処理状況に関する情報が格納されるプロセッサ情報格納領域と、を前記複数の記憶デバイスの記憶領域を用いて作成する。

【0011】

前記記憶デバイス制御部は、前記複数の第一のチャネル制御部内の前記第一のプロセッサの指示に応じて、前記プロセッサ情報格納領域に格納された情報を、前記複数の記憶デバイスの記憶領域を用いて作成されたプロセッサ情報バックアップ用の格納領域に対してコピーするように制御する。

【0012】

また、本発明のディスクアレイ装置において、前記複数の第一のチャネル制御部内の前記第一のプロセッサは、前記第一のプロセッサが設けられている前記第一のチャネル制御部内の前記第二のプロセッサに対して、前記第一のプロセッサの処理状況に関する情報を前記プロセッサ情報格納領域に格納するように指示する。この場合、前記第一のプロセッサが設けられている前記第一のチャネル制御部内の前記第二のプロセッサは、前記第一のチャネル制御部の指示に応じて、前記第一のプロセッサの処理状況に関する情報を、前記プロセッサ情報格納領域に格納するように制御する。

【0013】

また、本発明のディスクアレイ装置において、前記複数の第一のチャネル制御部内の第二のプロセッサは、前記第二のプロセッサが設けられている前記第一のチャネル制御部内の前記第一のプロセッサの要求に従って、前記ブロックレベルのデータを前記キャッシュメモリに保存するとともに、前記ブロックレベルのデータを前記キャッシュメモリに保存したことを表す情報を前記共有メモリに格納する。この場合、共有メモリは、前記複数の第一のチャネル制御部内の第二のプロセッサの制御のもとに、前記ブロックレベルのデータが前記キャッシュメモリに保存されたことを表す情報が格納されるものである。

【0014】

また、本発明のディスクアレイ装置において、前記複数の第一のチャネル制御部内の前記第一のプロセッサは、前記記憶デバイス制御部に対して、前記プロセッサ情報格納領域に格納された情報を、前記プロセッサ情報バックアップ用の格納領域に対してコピーするように指示するものである。この場合、前記記憶デバイス制御部は、前記第一のプロセッサの指示に応じて、コピー処理を制御する。

【0015】

また、本発明のディスクアレイ装置において、前記複数の第一のチャネル制御部内の前記第一のプロセッサは、前記プロセッサ情報格納領域に格納された情報の読み出し又は書き込みが不可能になった場合、前記プロセッサ情報バックアップ用の格納領域に格納された情報を読み出し又は書き込むことにより、処理を継続する。

【0016】

また、本発明のディスクアレイ装置において、前記複数の第一のチャネル制御部は、複数のクラスタグループに分類されるものである。前記プロセッサ情報格納領域は、複数のプロセッサ情報格納部分を有するものである。前記複数のクラスタグループは、前記複数のプロセッサ情報格納部分のうちの各々異なる部分を割当てられるものである。

【0017】

また、本発明のディスクアレイ装置において、前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部は、前記複数のプロセッサ情報格納部分のうちの第一のプロセッサ情報格納部分に対して、前記第一のプロセッサによって相互にやり取りされるプロセッサ間の処理状況に関する情報を格納する。この場合、前記複数のクラスタグループのうちの第二のクラスタグループに含まれる前記複数の第一のチャンネル制御部は、前記複数のプロセッサ情報格納部分のうちの第二のプロセッサ情報格納部分に対して、前記第一のプロセッサによって相互にやり取りされるプロセッサ間の処理状況に関する情報を格納する。

【0018】

また、本発明のディスクアレイ装置において、前記第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部の前記第一のプロセッサは、前記第一のプロセッサ情報格納部分に格納された情報の複製の作成を、前記記憶デバイス制御部に対して指示する。この場合、前記記憶デバイス制御部は、前記第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部の前記第一のプロセッサの指示に応じて、前記プロセッサ情報バックアップ用の格納領域に含まれる第一のバックアップ領域に対して、前記第一のプロセッサ情報格納部分に格納された情報の複製を格納する。

【0019】

また、本発明のディスクアレイ装置において、前記第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部の前記第一のプロセッサは、前記第一のプロセッサ情報格納部分及び前記第二のプロセッサ情報格納部分に格納された情報の複製の作成を、前記記憶デバイス制御部に対して指示する。この場合、前記記憶デバイス制御部は、前記第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部の前記第一のプロセッサの指示に応じて、前記プロセッサ情報バックアップ用の格納領域に含まれる第一のバックアップ領域及び第二のバックアップ領域に対して、前記第一のプロセッサ情報格納部分及び前記第二のプロセッサ情報格納部分に格納された情報の複製を格納する。

【0020】

また、本発明のディスクアレイ装置において、前記複数の第一のチャンネル制御部及び前記記憶デバイス制御部に関する情報の取得に利用される管理端末と、を有する。この場合、前記記憶デバイス制御部は、前記管理端末からの指示に応じて、前記プロセッサ情報バックアップ用の格納領域に含まれる第一のバックアップ領域及び第二のバックアップ領域に対して、前記第一のプロセッサ情報格納部分及び前記第二のプロセッサ情報格納部分に格納された情報の複製を格納する。

【0021】

また、本発明のディスクアレイ装置において、前記第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部の前記第一のプロセッサは、前記第一のプロセッサ情報格納部分に格納された情報の読み出し又は書き込みが不可能になった場合、前記第一のバックアップ領域に格納された情報を読み出し又は書き込むことにより、処理を継続する。

【0022】

また、本発明のディスクアレイ装置は、以下の構成を有する。

【0023】

ディスクアレイ装置は、データを格納する複数の記憶デバイスと、前記複数の記憶デバイスに対するデータの格納を制御する記憶デバイス制御部と、前記記憶デバイス制御部に接続される接続部と、複数の第一のチャンネル制御部と、前記複数の第一のチャンネル制御部及び前記記憶デバイス制御部によってやり取りされる制御情報が格納される共有メモリと、前記複数の第一のチャンネル制御部と前記記憶デバイス制御部との間でやり取りされるデータを一時的に保存するキャッシュメモリと、を有する。

【0024】

複数の第一のチャンネル制御部は、自ディスクアレイ装置の外部のローカルエリアネットワークを介して受けたファイルレベルのデータをブロックレベルのデータに変換して、前

記複数の記憶デバイスへの格納を要求する第一のプロセッサと、前記第一のプロセッサからの要求に応じて前記接続部及び前記記憶デバイス制御部を介して前記複数の記憶デバイスへ前記ブロックレベルのデータを転送する第二のプロセッサとを有し、前記接続部及び前記ローカルエリアネットワークに接続される。

【0025】

前記複数の第一のチャンネル制御部内の前記第二のプロセッサは、前記ブロックレベルのデータが格納される複数の記憶領域と、複数の前記第一のプロセッサによって相互にやり取りされるプロセッサ間の処理状況に関する情報が格納されるプロセッサ情報格納領域と、複数の前記第一のプロセッサ上で動作するソフトウェアプログラムが格納されるソフトウェアプログラム格納領域と、を前記複数の記憶デバイスの記憶領域を用いて作成する。

【0026】

前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記第一のプロセッサが設けられている前記第一のチャンネル制御部内の前記第二のプロセッサの制御に応じて、前記ソフトウェアプログラム格納領域に格納されているソフトウェアプログラムを取得して、動作する。

【0027】

また、本発明のディスクアレイ装置において、前記複数の第一のチャンネル制御部内の前記第一のプロセッサで動作するソフトウェアプログラムは、前記記憶デバイス制御部に対して、前記プロセッサ情報格納領域に格納された情報を、前記プロセッサ情報バックアップ用の格納領域に対してコピーするように指示する。この場合、前記記憶デバイス制御部は、前記複数の第一のチャンネル制御部内の前記第一のプロセッサの指示に応じて、前記プロセッサ情報格納領域に格納された情報を、前記複数の記憶デバイスの記憶領域を用いて作成されたプロセッサ情報バックアップ用の格納領域に対してコピーするように制御する。

【0028】

また、本発明のディスクアレイ装置において、前記複数の第一のチャンネル制御部は、複数のクラスタグループに分類されるものである。前記プロセッサ情報格納領域は、複数のプロセッサ情報格納部分を有するものである。前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部で動作するソフトウェアプログラムの各々は、相互に、前記複数のプロセッサ情報格納部分のうちの第一のプロセッサ情報格納部分に対して、プロセッサ間の処理状況に関する情報を格納しながら動作する。

【0029】

また、本発明のディスクアレイ装置において、前記複数のプロセッサ情報格納部分に格納された情報は、前記複数のクラスタグループ毎に、前記複数のプロセッサ情報格納部分に対応する複数のバックアップ領域に複製されるものである。

【0030】

また、本発明のディスクアレイ装置において、前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記第一のプロセッサが設けられている前記第一のチャンネル制御部内の前記第二のプロセッサを介して、前記記憶デバイス制御部に対して、ブロック単位で前記複製を実行するように指示する。この場合、前記記憶デバイス制御部は、前記第一のプロセッサの指示に応じて、前記複数のプロセッサ情報格納部分のうちの第一のプロセッサ情報格納部分に格納された情報を、ブロック単位で、前記複数のバックアップ領域のうちの第一のバックアップ領域に複製する。

【0031】

また、本発明のディスクアレイ装置において、前記ローカルエリアネットワークには、端末が設けられる。この場合、前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記端末からの指示に応じて、前記第一のプロセッサが設けられている前記第一のチャンネル制御部

内の前記第二のプロセッサを介して、前記記憶デバイス制御部に対して、ブロック単位で前記複製を実行するように指示する。

【0032】

また、本発明のディスクアレイ装置において、前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、一定の時間間隔毎に、前記第一のプロセッサが設けられている前記第一のチャンネル制御部内の前記第二のプロセッサを介して、前記記憶デバイス制御部に対して、ブロック単位で前記複製を実行するように指示する。

【0033】

また、本発明のディスクアレイ装置において、前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記第一のプロセッサが設けられている前記第一のチャンネル制御部内の前記第二のプロセッサを介して、前記記憶デバイス制御部の負荷状態を取得し、前記記憶デバイス制御部の負荷状態に応じて、前記記憶デバイス制御部に対して、ブロック単位で前記複製を実行するように指示する。

【0034】

また、本発明のディスクアレイ装置において、前記複数のクラスタグループのうちの第一のクラスタグループに含まれる前記複数の第一のチャンネル制御部内の前記第一のプロセッサは、前記複数のプロセッサ情報格納部分のうちの第一のプロセッサ情報格納部分に格納された情報へのアクセスが不可能になった場合に、前記第一のバックアップ領域に格納された情報を用いて処理を実行する。前記第一のプロセッサは、前記第一のプロセッサ情報格納部分が新たに形成された場合に、前記第一のバックアップ領域に格納された情報を前記新たに形成された第一のプロセッサ情報格納部分に複製して、前記新たに形成された第一のプロセッサ情報格納部分に格納された情報を用いて、処理を実行する。

【発明の効果】

【0035】

本発明によれば、複数の異種ネットワークに接続可能なように全く新しく発明された記憶装置システムを提供することができ、さらに、かかる記憶装置システムを発明するにあたり必要とされる記憶デバイス制御装置のシステム領域の複製を制御する方法をも提供することができる。

【発明を実施するための最良の形態】

【0036】

以下、本発明の実施の形態について図面を用いて詳細に説明する。

【0037】

まず、本実施の形態に係る記憶装置システムの全体構成を示すブロック図を図1に示す。

(全体構成例)

記憶装置システム600は、記憶デバイス制御装置100と記憶デバイス300とを備えている。記憶デバイス制御装置100は、情報処理装置200から受信したコマンドに従って記憶デバイス300に対する制御を行う。例えば情報処理装置200からデータの入出力要求を受信して、記憶デバイス300に記憶されているデータの入出力のための処理を行う。データは、記憶デバイス300が備えるディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域である論理ボリューム(Logical Unit) (以下、LUと記す) に記憶されている。また、記憶デバイス制御装置100は、情報処理装置200との間で、記憶装置システム600を管理するための各種コマンドの授受も行う。

【0038】

情報処理装置200はCPU (Central Processing Unit) やメモリを備えたコンピュータである。情報処理装置200が備えるCPUにより各種プログラムが実行されることによりさまざまな機能が実現される。情報処理装置200は、例えばパーソナルコンピュータや

ワークステーションであることもあるし、メインフレームコンピュータであることもある。

【0039】

図1において、情報処理装置1乃至3(200)は、LAN(Local Area Network)400を介して記憶デバイス制御装置100と接続されている。LAN400は、インターネットとすることもできるし、専用のネットワークとすることもできる。LAN400を介して行われる情報処理装置1乃至3(200)と記憶デバイス制御装置100との間の通信は、例えばTCP/IPプロトコルに従って行われる。情報処理装置1乃至3(200)からは、記憶装置システム600に対して、ファイル名指定によるデータアクセス要求(ファイル単位でのデータ入出力要求。以下、ファイルアクセス要求と記す)が送信される。

【0040】

LAN400にはバックアップデバイス910が接続されている。バックアップデバイス910は具体的にはMOやCD-R、DVD-RAMなどのディスク系デバイス、DATテープ、カセットテープ、オープンテープ、カートリッジテープなどのテープ系デバイスである。バックアップデバイス910は、LAN400を介して記憶デバイス制御装置100との間で通信を行うことにより、記憶デバイス300に記憶されているデータのバックアップデータを記憶する。またバックアップデバイス910は情報処理装置1(200)と接続されるようにすることもできる。この場合は情報処理装置1(200)を介して記憶デバイス300に記憶されているデータのバックアップデータを取得するようにする。

【0041】

記憶デバイス制御装置100は、チャンネル制御部1乃至4(110)を備える。記憶デバイス制御装置100は、チャンネル制御部1乃至4(110)によりLAN400を介して情報処理装置1乃至3(200)からのファイルアクセス要求を個々に受け付ける。すなわち、チャンネル制御部1乃至4(110)には、個々にLAN400上のネットワークアドレス(例えば、IPアドレス)が割り当てられていてそれぞれが個別にNASとして振る舞い、個々のNASがあたかも独立したNASが存在するように、NASとしてのサービスを情報処理装置1乃至3(200)に提供することができる。以下、チャンネル制御部1乃至4(110)をCHNと記す。このように1台の記憶装置システム600に個別にNASとしてのサービスを提供するチャンネル制御部1乃至4(110)を備えるように構成したことで、従来、独立したコンピュータで個々に運用されていたNASサーバが一台の記憶システム600に集約される。そして、これにより記憶装置システム600の統括的な管理が可能となり、各種設定・制御や生涯管理、バージョン管理といった保守業務の効率化が図られる。

【0042】

なお、本実施の形態に係る記憶デバイス制御装置100のチャンネル制御部1乃至4(110)は、後述するように、一体的にユニット化された回路基板上に形成されたハードウェアおよびこのハードウェアにより実行されるオペレーティングシステム(以下、OSと記す)やこのOS上で動作するアプリケーションプログラム、あるいはこのハードウェアにより実行される実行可能オブジェクトコードなどのソフトウェアにより実現される。このように本実施例の記憶装置システム600では、従来ハードウェアの一部として実装されてきた機能がソフトウェアにより実現されている。このため、本実施例の記憶装置システム600では柔軟性に富んだシステム運用が可能となり、多様で変化の激しいユーザーニーズによりきめ細かなサービスを提供することが可能となる。

【0043】

情報処理装置3乃至4(200)はSAN(Storage Area Network)500を介して記憶デバイス制御装置100と接続されている。SAN500は、記憶デバイス300が提供する記憶領域におけるデータの管理単位であるブロックを単位として情報処理装置3乃至4(200)との間でデータの授受を行うためのネットワークである。SAN500を介して行われる情報処理装置3乃至4(200)と記憶デバイス制御装置100との間の通信は、一般にファイバチャネルプロトコルに従って行われる。情報処理装置3乃至4からは、記憶装置システム600に対して、ファイバチャネルプロトコルに従ってブロック単位の

データアクセス要求（以下、ブロックアクセス要求と記す）が送信される。

【0 0 4 4】

SAN 5 0 0 には SAN 対応のバックアップデバイス 9 0 0 が接続されている。SAN 対応バックアップデバイス 9 0 0 は、SAN 5 0 0 を介して記憶デバイス制御装置 1 0 0 との間で通信を行うことにより、記憶デバイス 3 0 0 に記憶されているデータのバックアップデータを記憶する。

【0 0 4 5】

記憶デバイス制御装置 5 (2 0 0) は、LAN 4 0 0 や SAN 500 等のネットワークを介さずに記憶デバイス制御装置 1 0 0 と接続されている。情報処理装置 5 (2 0 0) としては例えばメインフレームコンピュータとすることができる。情報処理装置 5 (2 0 0) と記憶デバイス制御装置 1 0 0 との間の通信は、例えば FICON (Fibre Connection) (登録商標) や ESCON (Enterprise System Connection) (登録商標)、ACONARC (Advanced Connection Architecture) (登録商標)、FIBARC (Fibre Connection Architecture) (登録商標) などの通信プロトコルに従って行われる。情報処理装置 5 (2 0 0) からは、記憶装置システム 6 0 0 に対して、これらの通信プロトコルに従ってブロックアクセス要求が送信される。

【0 0 4 6】

記憶デバイス制御装置 1 0 0 は、チャンネル制御部 7 乃至 8 (1 1 0) により情報処理装置 5 (2 0 0) との間で通信を行う。以下、チャンネル制御部 7 乃至 8 (1 1 0) を CHA と記す。

【0 0 4 7】

SAN 5 0 0 には記憶装置システム 6 0 0 の設置場所（プライマリサイト）とは遠隔した場所（セカンダリサイト）に設置される他の記憶装置システム 6 1 0 が接続している。記憶装置システム 6 1 0 は、後述するレプリケーション又はリモートコピーの機能におけるデータの複製先の装置として利用される。なお、記憶装置システム 6 1 0 は SAN 5 0 0 以外にも ATM などの通信回線により記憶装置システム 6 0 0 に接続していることもある。この場合には例えばチャンネル制御部 1 1 0 として上記通信回線を利用するためのインタフェース（チャンネルエクステンダ）を備えるチャンネル制御部 1 1 0 が採用される。

（記憶デバイス）

記憶デバイス 3 0 0 は、多数のディスクドライブ（物理ディスク）を備えており、情報処理装置 2 0 0 に対して記憶領域を提供する。データは、ディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域である LU に記憶されている。ディスクドライブとしては、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等さまざまなものを用いることができる。なお、記憶デバイス 3 0 0 は例えば複数のディスクドライブによりディスクアレイを構成するようにすることもできる。この場合、情報処理装置 2 0 0 に対して提供される記憶領域は、RAID により管理された複数のディスクドライブにより提供されるようにすることもできる。

【0 0 4 8】

記憶デバイス制御装置 1 0 0 と記憶デバイス 3 0 0 との間は図 1 のように直接に接続される形態とすることもできるし、ネットワークを介して接続するようにすることもできる。さらに記憶デバイス 3 0 0 は記憶デバイス制御装置 1 0 0 と一体として構成されることもできる。

【0 0 4 9】

記憶デバイス 3 0 0 に設定される LU には、情報処理装置 2 0 0 からアクセス可能なユーザ LU や、チャンネル制御部 1 1 0 の制御のために使用されるシステム LU 等がある。システム LU には CHN 1 1 0 で実行される OS も格納される。また各 LU にはチャンネル制御部 1 1 0 が対応付けられている。これによりチャンネル制御部 1 1 0 ごとにアクセス可能な LU が割り当てられている。また上記対応付けは、複数のチャンネル制御部 1 1 0 で一つの LU を共有するようにすることもできる。なお以下において、ユーザ LU やシステム LU をそれぞれユーザディスク、システムディスク等とも記す。

(記憶デバイス制御装置)

記憶デバイス制御装置 100 は、チャンネル制御部 110、共有メモリ 120、キャッシュメモリ 130、ディスク制御部 140、管理端末 160 及び接続部 150 を備える。

【0050】

チャンネル制御部 110 は、情報処理装置 200 との間で通信を行うための通信インタフェースを備え、情報処理装置 200 との間でデータ入出力コマンド等を授受する機能を備える。例えば CHN 110 は情報処理装置 1 乃至 3 (200) からのファイルアクセス要求を受け付ける。これによる記憶装置システム 600 は NAS としてのサービスを情報処理装置 1 乃至 3 (200) に提供することができる。また CHF 110 は情報処理装置 3 乃至 4 (200) からのファイバチャネルプロトコルに従ったブロックアクセス要求を受け付ける。これにより記憶装置システム 600 は高速アクセス可能なデータ記憶サービスを情報処理装置 3 乃至 4 (200) に対して提供することができる。また CHA 110 は情報処理装置 5 (200) からの FICON や ESCON、ACONARC、FIBER C 等のプロトコルに従ったブロックアクセス要求を受け付ける。これにより記憶装置システム 600 は情報処理装置 5 (200) のようなメインフレームコンピュータに対してもデータ記憶サービスを提供することができる。

【0051】

各チャンネル制御部 110 は、管理端末 160 とともに内部 LAN 151 等の通信網で接続されている。これによりチャンネル制御部 110 に実行させるマイクロプログラム等を管理端末 160 から送信しインストールすることが可能となっている。チャンネル制御部 110 の構成については後述する。

【0052】

接続部 150 はチャンネル制御部 110、共有メモリ 120、キャッシュメモリ 130 及びディスク制御部 140 と接続されている。チャンネル制御部 110、共有メモリ 120、キャッシュメモリ 130 及びディスク制御部 140 間でのデータやコマンドの授受は、接続部 150 を介することにより行われる。接続部 150 は、例えば高速スイッチングによりデータ伝送を行う超高速クロスバススイッチなどのスイッチ、又はバス等で構成される。チャンネル制御部 110 同士がスイッチで接続されていることで、個々のコンピュータ上で動作する NAS サーバが LAN を通じて接続する従来の構成に比べてチャンネル制御部 110 間の通信パフォーマンスが大幅に向上している。また、これにより高速なファイル共有機能や高速フェイルオーバーなどが可能となる。

【0053】

共有メモリ 120 およびキャッシュメモリ 130 は、チャンネル制御部 110、ディスク制御部 140 により共有される記憶メモリである。共有メモリ 120 は主に制御情報やコマンド等を記憶する為に利用されるのに対し、キャッシュメモリ 130 は主にデータを記憶するために利用される。

【0054】

例えば、あるチャンネル制御部 110 が情報処理装置 200 から受信したデータ入出力コマンドが書き込みコマンドであった場合には、当該チャンネル制御部 110 は、書き込みコマンドを共有メモリ 120 に書き込むとともに、情報処理装置 200 から受信した書き込みデータをキャッシュメモリ 130 に書き込む。一方、ディスク制御部 140 は共有メモリ 120 を監視しており、共有メモリ 120 に書き込みコマンドが書き込まれたことを検出すると、当該コマンドに従ってキャッシュメモリ 130 から書き込みデータを読み出して記憶デバイス 300 に書き込む。

【0055】

また、例えば、あるチャンネル制御部 110 が情報処理装置 200 から受信したデータ入出力コマンドが読み出しコマンドであった場合には、当該チャンネル制御部 110 は、読み出しコマンドを共有メモリ 120 に書き込むとともに、情報処理装置 200 から読み出しコマンドによって要求されたデータをキャッシュメモリ 130 から読み出す。仮に読み出しコマンドによって要求されたデータがキャッシュメモリ 130 に書き込まれていなかっ

た場合、チャンネル制御部 1 1 0 又はディスク制御部 1 4 0 は、読み出しコマンドによって要求されたデータを記憶デバイス 3 0 0 から読み出して、キャッシュメモリ 1 3 0 に書き込む。

【0 0 5 6】

なお、上記の本実施の形態においては、共有メモリ 1 2 0 及びキャッシュメモリ 1 3 0 がチャンネル制御部 1 1 0 及びディスク制御部 1 4 0 に対して独立に設けられていることについて記載されているが、本実施の形態はこの場合に限られるものでなく、共有メモリ 1 2 0 又はキャッシュメモリ 1 3 0 がチャンネル制御部 1 1 0 及びディスク制御部 1 4 0 の各々に分散されて設けられることも好ましい。この場合、接続部 1 5 0 は、分散された共有メモリ又はキャッシュメモリを有するチャンネル制御部 1 1 0 及びディスク制御部 1 4 0 を相互に接続させることになる。

【0 0 5 7】

ディスク制御部 1 4 0 は、記憶デバイス 3 0 0 の制御を行う。例えば上述のように、チャンネル制御部 1 1 0 が情報処理装置 2 0 0 から受信したデータ書き込みコマンドに従って記憶デバイス 3 0 0 へデータの書き込みを行う。また、チャンネル制御部 1 1 0 により送信された論理アドレス指定による L U へのデータアクセス要求を、物理アドレス指定による物理ディスクへのデータアクセス要求に変換する。記憶デバイス 3 0 0 における物理ディスクが R A I D により管理されている場合には、R A I D 構成に従ったデータのアクセスを行う。またディスク制御部 1 4 0 は、記憶デバイス 3 0 0 に記憶されたデータの複製管理の制御やバックアップ制御を行う。さらにディスク制御部 1 4 0 は、災害発生時のデータ消失防止（ディザスタリカバリ）などを目的として、プライマリサイトの記憶装置システム 6 0 0 のデータの複製をセカンダリサイトに設置された他の記憶装置システム 6 1 0 にも記憶する制御（レプリケーション機能、またはリモートコピー機能）なども行う。

【0 0 5 8】

各ディスク制御部 1 4 0 は管理端末 1 6 0 とともに内部 L A N 1 5 1 等の通信網で接続されており、相互に通信を行うことが可能である。これにより、ディスク制御部 1 4 0 に実行させるマイクロプログラム等を管理端末 1 6 0 から送信しインストールすることが可能となっている。ディスク制御部 1 4 0 の構成については後述する。

（管理端末）

管理端末 1 6 0 は記憶装置システム 6 0 0 を保守・管理するためのコンピュータである。管理端末 1 6 0 を操作することにより、例えば記憶デバイス 3 0 0 内の物理ディスク構成の設定や、L U の設定、チャンネル制御部 1 1 0 において実行されるマイクロプログラムのインストール等を行うことができる。ここで、記憶デバイス 3 0 0 内の物理ディスク構成の設定としては、例えば物理ディスクの増設や減設、R A I D 構成の変更（R A I D 1 から R A I D 5 への変更等）等を行うことができる。さらに管理端末 1 6 0 からは、記憶装置システム 6 0 0 の動作状態の確認や故障部位の特定、チャンネル制御部 1 1 0 で実行される OS のインストール等の作業を行うこともできる。また管理端末 1 6 0 は L A N や電話回線等で外部保守センタと接続されており、管理端末 1 6 0 を利用して記憶装置システム 6 0 0 の障害監視を行ったり、障害が発生した場合に迅速に対応することも可能である。障害の発生は例えば OS やアプリケーションプログラム、ドライバソフトウェアなどから通知される。この通知は H T T P プロトコルや S N M P （Simple Network Management Protocol）、電子メールなどにより行われる。これらの設定や制御は、管理端末 1 6 0 で動作する W e b サーバが提供する W e b ページをユーザインタフェースとしてオペレータなどにより行われる。オペレータ等は、管理端末 1 6 0 を操作して障害監視する対象や内容の設定、障害通知先の設定などを行うこともできる。

【0 0 5 9】

管理端末 1 6 0 は記憶デバイス制御装置 1 0 0 に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。また管理端末 1 6 0 は、記憶デバイス制御装置 1 0 0 及び記憶デバイス 3 0 0 の保守・管理を専用に行うコンピュータとすることもできるし、汎用のコンピュータに保守・管理機能を持たせたものとすることもできる

【0060】

管理端末160の構成を示すブロック図を図2に示す。

【0061】

管理端末160は、CPU161、メモリ162、ポート163、記録媒体読み取り装置164、入力装置165、出力装置166及び記憶装置168を備える。

【0062】

CPU161は、管理端末160の全体の制御を司るもので、メモリ162に格納されたプログラム162cを実行することにより上記Webサーバとしての機能等を実現する。メモリ162には、物理ディスク管理テーブル162aとLU管理テーブル162bとプログラム162cとが記憶されている。

【0063】

物理ディスク管理テーブル162aは、記憶デバイス300に備えられる物理ディスク（ディスクドライブ）を管理するためのテーブルである。物理ディスク管理テーブル162aを図3に示す。図3においては、記憶デバイス300が備える多数の物理ディスクのうち、ディスク番号#001乃至#006までが示されている。それぞれの物理ディスクに対して、容量、RAID構成、使用状況が示されている。

【0064】

LU管理テーブル162bは、上記物理ディスク上に論理的に設定されるLUを管理するためのテーブルである。LU管理テーブル162bを図4に示す。図4においては、記憶デバイス300上に設定される多数のLUのうち、LU番号#1乃至#3までが示されている。それぞれのLUに対して、物理ディスク番号、容量、RAID構成が示されている。

【0065】

記憶媒体読取装置164は、記録媒体167に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ162や記憶装置168に格納される。従って、例えば記録媒体167に記録されたプログラム162cを、記録媒体読取装置164を用いて記録媒体167から読み取って、目盛り162や記憶装置168に格納するようにすることができる。記録媒体167としてはフレキシブルディスクやCD-ROM、半導体メモリ等を用いることができる。記録媒体読取装置162は管理端末160に内蔵されている形態とすることもできる。記憶装置168は、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等である。入力装置165は、オペレータ等による管理端末160へのデータ入力等のために用いられる。入力装置165としては例えばキーボードやマウス等が用いられる。出力装置166は、情報を外部に出力するための装置である。出力装置166としては例えばディスプレイやプリンタ等が用いられる。ポート163は内部LAN151に接続されており、これにより管理端末160はチャネル制御部110やディスク制御部140等と通信を行うことができる。またポート163は、LAN400に接続するようにすることもできるし、電話回線に接続するようにすることもできる。

（外観図）

次に、本実施の形態に係る記憶装置システム600の外観構成を図5に示す。

また、記憶デバイス制御装置100の外観構成を図6に示す。

【0066】

図5に示すように、本実施の形態に係る記憶装置システム600は記憶デバイス制御装置100および記憶デバイス300がそれぞれの筐体に収められた形態をしている。記憶デバイス制御装置100の筐体の両側に記憶デバイス300の筐体が配置されている。

【0067】

記憶デバイス制御装置100は、正面中央部に管理端末160が備えられている。管理端末160はカバーで覆われており、図6に示すようにカバーを開けることにより管理端末160を使用することができる。なお図6に示した管理端末160はいわゆるノート型

パーソナルコンピュータの形態をしているが、どのような形態とすることも可能である。

【0068】

管理端末160の下部には、チャンネル制御部110を装着するためのスロットが設けられている。各スロットにはチャンネル制御部110のボードが装着される。本実施の形態に係る記憶装置システム600においては、例えばスロットは8つあり、図5および図6には8つのスロットにチャンネル制御部110を装着するためのガイドレールが設けられている。ガイドレールに沿ってチャンネル制御部110をスロットに挿入することにより、チャンネル制御部110を記憶デバイス制御装置100に装着することができる。また各スロットに装着されたチャンネル制御部110は、ガイドレールに沿って手前方向に引き抜くことにより取り外すことができる。また各スロットの奥手方向正面部には、各チャンネル制御部110を記憶デバイス制御装置100と電氣的に接続するためのコネクタが設けられている。チャンネル制御部110には、CHN、CHF、CHAがあるが、いずれのチャンネル制御部110もサイズやコネクタの位置、コネクタのピン配列等に互換性をもたせているため、8つのスロットにはいずれのチャンネル制御部110も装着することが可能である。従って、例えば8つのスロット全てにCHN110を装着するようにすることもできる。また例えば図1に示したように、4枚のCHN110と、2枚のCHF110と、2枚のCHA110とを装着するようにすることもできる。チャンネル制御部110を装着しないスロットを設けることもできる。

【0069】

なお、上述したように、チャンネル制御部110は上記各スロットに装着可能なボード、すなわち同一のユニットに形成された一つのユニットとして提供されるが、上記同一のユニットは複数枚数の基板から構成されているようにすることもできる。つまり、複数枚数の基板から構成されていても、各基板が相互に接続されて一つのユニットとして構成され、記憶デバイス制御装置100のスロットに対して一体的に装着できる場合は、同一の回路基板の概念に含まれる。

【0070】

ディスク制御部140や共有メモリ120等の、記憶デバイス制御装置100を構成する他の装置については図5および図6には示されていないが、記憶デバイス制御装置100の背面側面に装着されている。

【0071】

また記憶デバイス制御装置100には、チャンネル制御部110とらおから発生する熱を放出するためのファン170が設けられている。ファン170は記憶デバイス制御装置100の上面部に設けられるほか、チャンネル制御部110用スロットの上部にも設けられている。

【0072】

ところで、筐体に収容されて構成される記憶デバイス制御装置100および記憶デバイス300としては、例えばSAN製品として製品化されている従来構成の装置を利用することができる。特に上記のようにCHNのコネクタ形状を従来構成の筐体に設けられているスロットにそのまま装着できる形状とすることで従来構成の装置をより簡単に利用することができる。つまり、本実施例の記憶装置システム600は、既存の製品を利用することで容易に構築することができる。

【0073】

さらに、本実施の形態によれば、記憶装置システム600内にCHN110、CHF110、CHA110を混在させて装着させることにより、異種ネットワークに接続される記憶装置システムを実現できる。具体的には、記憶装置システム600は、CHN110を用いてLAN140に接続し、かつCHF110を用いてSAN500に接続するという、SAN-NAS統合記憶装置システムである。

(チャンネル制御部)

本実施の形態に係る記憶装置システム600は、上述の通りCHN110により情報処理装置1乃至3(200)からのファイルアクセス要求を受け付け、NASとしてのサー

ビスを情報処理装置 1 乃至 3 (200) に提供する。

【0074】

CHN110 のハードウェア構成を図 7 に示す。この図に示すように CHN110 のハードウェアは一つのユニットで構成される。以下、このユニットのことを NAS ボードと記す。NAS ボードは一枚もしくは複数枚の回路基板を含んで構成される。より具体的には、NAS ボードは、ネットワークインタフェース部 111、入出力制御部 114、ボード接続用コネクタ 116、通信コネクタ 117 及びファイルサーバ部 800 を備え、これらが同一のユニットに形成されて構成されている。さらに、入出力制御部 114 は、NVRAM (Non Volatile RAM) 115 及び I/O (Input/Output) プロセッサ 119 を有する。

【0075】

ネットワークインタフェース部 111 は、情報処理装置 200 との間で通信を行うための通信インタフェースを備えている。CHN110 の場合は、例えば TCP/IP プロトコルに従って情報処理装置 200 から送信されたファイルアクセス要求を受信する。通信コネクタ 117 は、情報処理装置 200 との間で通信を行うためのコネクタである。CHN110 の場合は、LAN400 に接続可能なコネクタであり、例えばイーサネット (登録商標) に対応している。

【0076】

ファイルサーバ部 800 は、CPU112、メモリ 113、BIOS (Basic Input/Output System) 801 及び NVRAM 804 を有する。CPU112 は、CHN110 を NAS ボードとして機能させるための制御を司る。CPU112 は、NFS 又は CIFS 等のファイル共有プロトコル及び TCP/IP の制御、ファイル指定されたファイルアクセス要求の解析、メモリ 113 内の制御情報へのファイル単位のデータと記憶デバイス 300 内の LU との変換テーブル (図示せず) を用いた相互変換、記憶デバイス 300 内の LU に対するデータ書き込み又は読み出し要求の生成、データ書き込み又は読み出し要求の I/O プロセッサ 119 への送信等処理する。BIOS 801 は、例えば CHN110 に電源が投入された際に、CPU112 を起動する過程で最初にメモリ 113 にロードされ実行されるソフトウェアであり、例えばフラッシュメモリなどの不揮発性の媒体に保存されて CHN110 上に実装されている。CPU112 は、BIOS 801 からメモリ 113 上に読み込まれたソフトウェアを実行することにより、CHN21 上の CPU112 が関係する部分の初期化、診断などを行うことができる。さらに、CPU112 は、BIOS 801 から I/O プロセッサ 119 にコマンドなどの指示を発行することにより、記憶デバイス 300 から所定のプログラム、例えば OS のブート部などをメモリ 113 に読み込むことができる。読み込まれた OS のブート部は、さらに記憶デバイス 300 に格納されている OS の主要部分をメモリ 113 に読み込む動作をし、これにより CPU112 上で OS が起動され、例えばファイルサーバとしての処理が実行できるようになる。また、ファイルサーバ部 800 は、PXE (Preboot eXecution Environment) などの規約にしたがうネットワークブートローダを格納する NVRAM 804 を実装し、後述するネットワークブートを行わせることも可能である。

【0077】

メモリ 113 にはさまざまなプログラムやデータが記憶される。例えば図 8 に示すメタデータ 730 やロックテーブル 720、また図 14 に示される NAS マネージャ 706 等の各種プログラムが記憶される。メタデータ 730 は、ファイルシステムが管理しているファイルに対応させて生成される情報である。メタデータ 730 には例えばファイルのデータが記憶されている LU 上のアドレスやデータサイズなど、ファイルの保管場所を特定するための情報が含まれる。メタデータ 730 にはファイルの容量、所有者、更新時刻等の情報が含まれることもある。また、メタデータ 730 はファイルだけでなくディレクトリに対応させて生成されることもある。メタデータ 730 の例を図 9 に示す。メタデータ 730 は記憶デバイス 300 上の各 LU にも記憶されている。

【0078】

ロックテーブル 720 は、情報処理装置 1 乃至 3 (200) からのファイルアクセスに対して排他制御を行うためのテーブルである。排他制御を行うことにより情報処理装置 1 乃至 3 (200) でファイルを共用することができる。ロックテーブル 720 を図 10 に示す。図 10 に示すようにロックテーブル 720 には、ファイルロックテーブル 721 と LU ロックテーブル 722 とがある。ファイルロックテーブル 721 は、ファイルごとにロックが掛けられているか否かを示すためのテーブルである。いずれかの情報処理装置 200 によりあるファイルがオープンされている場合に当該ファイルにロックが掛けられる。ロックが掛けられたファイルに対する他の情報処理装置 200 によるアクセスは禁止される。LU ロックテーブル 722 は、LU ごとにロックが掛けられているか否かを示すためのテーブルである。いずれかの情報処理装置 200 により、ある LU に対するアクセスが行われている場合に当該 LU にロックが掛けられる。ロックが掛けられた LU に対する他の情報処理装置 200 によるアクセスは禁止される。

【0079】

入出力制御部 114 は、ディスク制御部 140 キャッシュメモリ 130、共有メモリ 120 及び管理端末 160 との間でデータやコマンドの授受を行う。入出力制御部 114 は I/O プロセッサ 119 及び NVRAM 115 を備えている。I/O プロセッサ 119 は例えば 1 チップのマイコンで構成される。I/O プロセッサ 119 は、記憶デバイス 300 内の LU に対するデータ書き込み又は読み出し要求やデータの授受を制御し、CPU 112 とディスク制御部 140 との間の通信を中継する。NVRAM 115 は I/O プロセッサ 119 の制御を司るプログラムを格納する不揮発性メモリである。NVRAM 115 に記憶されるプログラムの内容は、管理端末 160 や、後述する NAS マネージャ 706 からの指示により書き込みや書き換えを行うことができる。

【0080】

図 11 は、CHN 110 上の CPU 112 と I/O プロセッサ 119 との通信経路について具体的に示す。I/O プロセッサ 119 と、CPU 112 は、CHN 110 上に実装された通信メモリ 802、ハードウェアレジスタ群 803 で物理的に接続されている。通信メモリ 802 およびハードウェアレジスタ群 803 は、それぞれ CPU 112 および I/O プロセッサ 119 のいずれからアクセスが可能である。ハードウェアレジスタ群 803 は、CPU 112 に対して電源を投入又は切断する回路に接続される。これにより、I/O プロセッサ 119 は、ハードウェアレジスタ群 803 にアクセスすることによって、ハードウェアレジスタ群 803 を介して CPU 112 の電源を操作することが可能となる。ハードウェアレジスタ群 803 は、必要に応じて、CPU 112 あるいは I/O プロセッサ 119 がハードウェアレジスタ群 803 にアクセスを行った際に、アクセス対象の相手先に割り込み信号などを生成して、アクセスが行われたことを通知する等の複数の機能を有する。これら複数の機能は、ハードウェアレジスタ群 803 を構成する各レジスタにそれぞれハードウェア的に割り当てられる。

【0081】

図 12 は、CPU 112 と I/O プロセッサ 119 とを、内部 LAN 151 によって接続しているハードウェア構成図である。このように、CPU 112 と I/O プロセッサ 119 とは、ともに内部 LAN 151 によっても接続されており、内部 LAN 151 を介して管理端末 160 との通信が可能である。これにより、例えば、CPU 112 は、NVRAM 804 に予め格納されているネットワークブートローダを実行することにより、管理端末 160 から起動用のソフトウェアをメモリ 113 にダウンロードし、起動用のソフトウェアを実行することができる。これによって例えば、管理端末 160 をサーバとし、CPU 112 をクライアントとするネットワークブートプロセスが実行される。なお、ネットワークブートは、例えば PXE などの規約に従い、クライアント上のネットワークブートローダと管理端末 160 上で動作するサーバとが、IP プロトコル、DHCP、TFTP、FTP などのプロトコルを組み合わせることにより、LAN 上の管理端末 160 に存在する OS のブートイメージを起動及び実行する方法である。

【0082】

図13は、ディスク制御部140のハードウェア構成を示すブロック図である。既に述べた通り、ディスク制御部は、記憶デバイス300に接続されるとともに接続部150を介してCHN112に接続され、ディスク制御部140独自で、又はCHN112によって制御されることにより、記憶デバイス300に対してデータの読み書きを行う。

【0083】

ディスク制御部140は、インタフェース部141、メモリ143、CPU142、NVRAM144及びボード接続用コネクタ145を備え、これらが一体的なユニットとして形成されている。

【0084】

インタフェース部141は、接続部150を介してチャネル制御部110等と通信を行うための通信インタフェース、記憶デバイス300と通信を行うための通信インタフェース、内部LAN151を介して管理端末160と通信を行うための通信インタフェースを備えている。

【0085】

CPU142は、ディスク制御部140全体の制御を司るとともに、チャネル制御部110や記憶デバイス300、管理端末160との間の通信を行う。メモリ143やNVRAM144に格納された各種プログラムを実行することにより本実施の形態に係るディスク制御部140の機能が実現される。ディスク制御部140により実現される機能としては、記憶デバイス300の制御やRAID制御、記憶デバイス300に記憶されたデータの複製管理やバックアップ制御、リモートコピー制御等である。

【0086】

NVRAM144はCPU142の制御を司るプログラムを格納する不揮発性メモリである。NVRAM144に記憶されるプログラムの内容は、管理端末160や、NASマネージャ706からの指示により書き込みや書き換えを行うことができる。

【0087】

またディスク制御部140はボード接続用コネクタ145を備えている。ボード接続用コネクタ145が記憶デバイス制御装置100側のコネクタと接続することにより、ディスク制御部140は、記憶デバイス制御装置100と電氣的に接続される。

(ソフトウェア構成図)

図14は、本実施の形態に係る記憶装置システム600におけるソフトウェア構成図である。既に述べたように、CHN110上には、CPU112およびI/Oプロセッサ119が存在する。CPU112およびI/Oプロセッサ119は、それぞれ1つずつであってもよいし、それぞれ複数存在してもよい。CPU112上では、OS701とNASマネージャ706等の多様なアプリケーションとが実行されることにより、CPU112はNASサーバとして動作する。I/Oプロセッサ119上では、コントローラとしてのマイクロプログラムが動作している。ディスク制御部140では、RAID制御部740がCPU142上で動作している。管理端末160の上では、CPU161がネットブートサーバ703として動作する。ネットブートサーバ703は、記録媒体167又は記憶装置168等から内部LAN151を介して、ミニカーネル704、OSイメージ705等をCHN110上のCPU112に転送する。ネットブートサーバ703は、例えば、DHCP (Dynamic Host Configuration Protocol) サーバなどを有し、CPU112、CPU161及びI/Oプロセッサ119にIPアドレス又はMACアドレスを割り当てる等して、管理端末160とCPU112、CPU161及びI/Oプロセッサ119との間の転送を行う。ネットブートを行うとき、例えば、CPU112は、クライアントとして、ネットブートサーバ703に対してDHCP要求及びファイル転送要求等を要求する。CPU112は、ネットブートの手順を経て、CPU112上でミニカーネル704を動作させることになる。最終的に、CPU112は、I/Oプロセッサ119を経由してOSイメージ705を記憶デバイス300にインストールさせる。

【0088】

なお、図14は、情報処理装置200のソフトウェア構成についても明示してある。情

報処理装置 2 0 0 は、N F S (Network File System) 7 1 1 を有するもの、又は C I F S (Common Internet File System) 7 1 3 を有するものが存在する。N F S 7 1 1 は、主に U N I X (登録商標) 系のオペレーティングシステム 7 1 4 によって用いられるファイル共有プロトコルであり、C I F S 7 1 3 は、主に W i n d o w s (登録商標) 系の O S 7 1 5 によって用いられるファイル共有プロトコルである。

(記憶装置システムのシステム領域)

図 1 5 は、情報処理装置 2 0 0 内部における、ソフトウェアや情報の格納領域を示している。C P U 1 1 2 のソフトウェアは、ネットワークインストールなどによって記憶デバイス 3 0 0 に格納されている。ここで記憶デバイスを L U (Logical Unit) 1 から L U 6 で表す。ここで C H N 1 の C P U 1 1 2 のソフトウェアが L U 1 に格納され、C H N 2 の C P U 1 1 2 のソフトウェアが L U 4 に格納されているものとする。L U 2 は C H N 1 の情報格納エリアとして予約されており、L U 5 は C H N 2 の情報格納エリアとして予約されている。また L U 3 は、C H N 1 の C P U 1 1 2 のソフトウェアと、C H N 2 の C P U 1 1 2 のソフトウェアが連携して動作するために必要な情報を格納する共有 L U である。さらに L U 6 は、L U 3 の情報をバックアップするための共有 L U バックアップ L U である。

【0089】

I O プロセッサ 1 1 9 は、C P U 1 1 2 からの指示を受けるか、管理端末 1 6 0 からの指示を受けることによって、共有 L U から共有 L U バックアップへのデータ転送を行うことができる。また、ディスク制御部 1 4 0 が共有 L U から共有 L U バックアップへのデータ転送を独自に行うこともできる。

【0090】

これにより、L U 3 の情報を使って例えば C H N 1 と C H N 2 との間でフェイルオーバーなどのオペレーションを行おうとした際、L U 3 が使用不能であった場合に、L U 3 の情報を使用する代わりに L U 6 の情報を使用することによって、問題なくフェイルオーバー動作を続行することができる。

【0091】

さらに、I O プロセッサ 1 1 9 は、C P U 1 1 2 からの指示を受けるか、管理端末 1 6 0 からの指示を受けることによって、L U 1 から L U 4 へ、L U 5 から L U 2 へと、互いに異なる C H N の情報格納エリアに対して、C P U 1 1 2 のソフトウェアをバックアップすることもできる。これにより例えば C H N 1 の C P U 1 1 2 のソフトウェア格納領域が使用不能になった場合に、L U 1 を保守員が交換した後に、C H N 1 のソフトウェアはインストールされていない状態に戻ってしまうが、C H N 2 の C P U から指示を行うことによって、L U 4 からソフトウェアを復元することができる。

【0092】

(記憶装置システムのデータアクセス方式)

一般に、オペレーティングシステムから見て記憶装置上のデータに対するアクセス方式には 2 種類がある。ひとつはファイルシステムを使用したアクセスであり、もう一つはファイルシステムを使用しないアクセスである。オペレーティングシステムは、システムコールと呼ばれる方法などにより、ファイルシステムを使用することなくデータへのアクセスを行うこともできる。ファイルシステムを使用しないアクセスは、記憶装置上のデータの位置を直接指定してアクセスすることになる。記憶装置上のデータの位置を直接指定してアクセスする場合、特段の処理を施さない場合は、複数のオペレーティングシステムから同時に同じ位置へのアクセスが発生した場合排他を行うことが出来ないため、オペレーティングシステム同士、あるいは複数のコントローラマイクロプログラム間、あるいは複数のディスク制御部間で何らかの手段でお互いに排他制御を行う必要がある。

【0093】

ファイルシステムとは、記憶装置上のデータに対する管理の方式、あるいは記憶装置上のデータを管理するソフトウェア、あるいは記憶装置上に格納された、記憶装置上のデータの管理情報などを指す一般的な用語である。通常オペレーティングシステムはファイル

システムを使用してデータにアクセスを行う。ファイルシステムのソフトウェアはデータの排他制御機能を通常実装しているため、複数のオペレーティングシステムが記憶装置上の同じ領域にあるデータを同時にアクセスしようとした場合でも、互いのファイルシステムの排他制御によってデータが保全される。ファイルシステムを用いてデータを管理する場合には、記憶装置上の領域に対してファイルシステムを定義し、定義したファイルシステムをオペレーティングシステムの管理用情報として登録したのちにファイルシステムに対してアクセス要求を行う必要がある。一般にファイルシステムの定義はファイルシステムの作成、ファイルシステムの登録はファイルシステムのマウントなどと呼ばれる。ファイルシステムは任意のタイミングで、オペレーティングシステムからの指示によりマウントを行ったり、マウントを取り消したりできる。マウントの取り消しはアンマウントという。

【0094】

通常、コントローラマイクロプログラムに対してI/Oなどの指示を直接行うのは、CPUで動作しているI/Oドライバである。オペレーティングシステムは通常ファイルシステムソフトウェアを使用して、I/Oドライバに対して、コントローラマイクロプログラムに向かって命令を発行するように要求する。この場合のアクセスはファイルシステムアクセスとなり、排他制御や、データの物理的格納位置についてはファイルシステムが管理する。また、オペレーティングシステムはファイルシステムソフトウェアを使用せずに、直接I/Oドライバに対して、コントローラマイクロプログラムに向かって命令を発行するように要求することもできる。この場合、データの位置や排他を管理するファイルシステムを使用しないため、オペレーティングシステムは何らかの手段でデータの位置を管理し、排他制御などを独自に行う必要がある。いずれの場合でも、コントローラマイクロプログラムから見ると、データの位置情報、転送サイズなどが指定された状態の要求がI/Oドライバから発行されてくることになる。これは、コントローラマイクロプログラムの立場からは、CPUからの要求がファイルシステムを使用したものか、そうでないのかの判断ができないことを意味する。

(CHNの動作方式)

本記憶装置において、高い可用性を保証するために、複数のCHNがひとつの組となって、互いに補完しあいながら動作することが可能である。これら複数のCHNによって作られる動作の単位をクラスタと呼ぶ。あるクラスタに属するCHNは、ユーザデータの格納されているLUへのパスを共有し、ユーザがクライアントからどのCHNに対して要求を発行しても、適切なLUへのアクセスができる状態になっている。ただしパスの定義は記憶装置システムのコントローラマイクロプログラムが認識する情報であるため、オペレーティングシステムが当該LUへアクセスするためには、通常はファイルシステムを使用しマウントを行う必要がある。パスが定義されていないとコントローラマイクロプログラムからオペレーティングシステムに対し当該LUの存在が伝達されないため、マウントも行えないが、パスが定義されていることによって、オペレーティングシステムがコントローラマイクロプログラムに対して問い合わせを行った際に、コントローラマイクロプログラムが当該LUの存在をオペレーティングシステムに伝達することができる。すなわち、オペレーティングシステムが当該LUにアクセスするためには、まずコントローラマイクロプログラムから当該LUへのアクセスパスを定義し、オペレーティングシステムがコントローラマイクロプログラムに対して使用可能なデバイスの問い合わせを行った際に当該LUの存在がコントローラマイクロプログラムから報告される必要があり、さらに、使用可能として報告されているデバイスの中から、オペレーティングシステムは最小で一つ、最大で報告されているデバイス全てについて、ファイルシステムの作成を行い、さらにそのファイルシステムをマウントすることによって実現される。ここでファイルシステムの作成とは、オペレーティングシステムが、当該デバイスに対して、ファイル名やディレクトリ名を指定してデータアクセスを行うことができるよう、ファイルやディレクトリの構造を定義し、その構造に対するアクセスのルールを定義し、これらの情報をシステム領域とデータ領域双方に記憶することを言う。本システムの場合システム領域はシステムLU

内に存在し、データ領域はユーザLU内に存在する。オペレーティングシステムは、このルールに従ってファイルやディレクトリ構造を操作することによって、データにアクセスする。このアクセス方式をファイルシステムアクセスという。

【0095】

(記憶装置システムの記憶装置に対するデータアクセス方法)

図16は、情報処理装置200において共有LUを複数のパーティションに分割し、それぞれを複製する様子を論理ブロック図で示している。

【0096】

共有LUは4個のパーティションに分割されており、同容量の共有LUバックアップLUも同じ容量ずつの4個のパーティションに分割されている。これらの設定は、CHNにオペレーティングシステムをインストールする際、管理端末160からの設定で共有LUおよび共有LUバックアップLUの初期化を行うよう指示することにより達成される。

(記憶装置システムのバックアップLU)

次に、実際に共有LUをバックアップする際の手順を、CHN1およびCHN5が共用している共有LUのパーティション部分のバックアップの例を用いて説明する。共有LUは共有LU311乃至314の4つのパーティションに分割される。パーティションの分割は、オペレーティングシステムの定義によって行われ、オペレーティングシステムからのアクセスによってのみ意味を持つ。共有LU311乃至314と、共有LUのバックアップ321乃至324は、それぞれCHN1およびCHN5からパスが定義されている。これは、CHN1のコントローラと、CHN5のコントローラから、共有LU311乃至314と、共有LUのバックアップ321乃至324にアクセスができることを意味する。この段階で、CHN1あるいはCHN5のオペレーティングシステムから、データブロックアクセス指示をCHN1、CHN5のコントローラに対して発行することによって、共有LU311乃至314と、共有LUのバックアップ321乃至324に対してデータの読み込み、あるいは書き出しの操作ができることを意味する。さらに、もしCHN1あるいはCHN5のオペレーティングシステムから共有LU311乃至314、共有LUのバックアップ321乃至324に対してファイルシステムを作成している場合、当該ファイルシステムをCHN1あるいはCHN5からマウントすれば、オペレーティングシステムは、共有LU311乃至314、共有LUのバックアップ321乃至324に対して、ファイルシステムを使用してデータの読み込み、書き出しを行うことができることになる。ここでは、ファイルシステムを使用したデータの読み込み、書き出しを行う例について述べる。CHN1、CHN2、CHN3、CHN5、CHN6、CHN7は、それぞれ自分が装着されているスロットの位置によって、アクセスすべきパーティションの位置を決定し、オペレーティングシステムはそれによって自分がアクセスすべき共有LU311乃至314と、共有LUのバックアップ321乃至324の場所を判定する。この例では、CHN1およびCHN5は、共有LU311と、共有LUのバックアップ321に対してアクセスすることが決定される。CHN1およびCHN5は、共有LU312乃至314と、共有LU322乃至324に対しては、オペレーティングシステムとしてはアクセスを行わない。

【0097】

(バックアップLUの定義)

まず初めに、各CHNは、それぞれ独自の共有LU311乃至314を持つ。共有LU311乃至314、共有LUのバックアップ321乃至324については、あらかじめ管理端末160からアクセスパスを定義しておく。CHNが複数存在する場合は、すべてのCHNに対して、これらの共有LU乃至311乃至314、共有LUのバックアップ321乃至324へのアクセスパスを定義しておく。

【0098】

次に、システムに一番初めにNASを導入する際に、最初にCHNを実装するタイミングで、オペレーティングシステムを管理端末160からネットワークインストールする。この際、ネットワークインストールの作業の一環として、インストールプログラムによっ

て、共有LU311乃至314、共有LUのバックアップ321乃至324の初期化を行う。またこの際に、共有LUを共有LU311乃至324、共有LUのバックアップを共有LUのバックアップ321乃至324というようにパーティションに分割し、それらの情報を共有LUに記憶させる。その作業が完了したのちに、オペレーティングシステムを、オペレーティングシステム用LUに管理端末160よりネットワークインストールする。

【0099】

以降、CHNを実装する際には、それぞれのCHNに対応するオペレーティングシステム用LUを順次初期化し、オペレーティングシステムをネットワークインストールしていくが、共有LU311乃至314、および共有LUのバックアップ321乃至324については、既に一度初期化してあるため以降は初期化しない。

【0100】

(共有LUの用途)

共有LU311に格納されるデータは、たとえばCHN間で処理の引継ぎを行う際の引き継ぎデータなどが格納される。CHN1は、処理を行っている際に、CHNのIPアドレス等のクライアントからのアクセスに必要な情報、クライアント情報や動作アプリケーション情報、オペレーションシステム上のサービスやデーモンの動作状態などの処理情報を共有LU311に格納する。もしCHN1がハードウェア障害やソフトウェア障害などで使用不能になった場合、これをハートビート機能などでCHN5が検知すると、共有LU311に格納されている上記の情報を元に、CHN1が行っていた処理を肩代わりし実行する。これにより、CHN1にアクセスしていたクライアントは、引き続きCHN5にアクセスすることにより処理を継続できる。この動作をフェイルオーバーという。

(共有LUのバックアップの必要性)

共有LU311は、通常RAIDシステムなどにより、物理的なハードディスクが1台故障しても動作が継続できるように設計されているが、予めRAIDとして用意された冗長度を超えた深刻度の故障が発生した場合などは、共有LU311は使用不能となる。この場合、さらにCHN1が故障した場合には、CHN5が処理を引き継ぐための情報を取得することができなくなる。このため共有LU311のデータを、共有LUのバックアップ321にコピーする必要がある。

(共有LUのバックアップ方式)

共有LUのバックアップには、例えばオペレーティングシステム上の汎用コマンドによって、コピーを行うことが考えられる。この場合、デバイスレベルでブロック単位でコピーを行うコマンドや、ファイル名を指定することによってファイル単位でコピーを行うコマンドなどがある。このコマンドを、ネットワーク上に存在するクライアントワークステーションやパーソナルコンピュータなどからオペレーティングシステムにログインし、端末を表示させて端末上で入力することにより、オペレーティングシステムに実行させることによってバックアップを実行する。これらのコマンドは、例えばオペレーティングシステムがUNIX（登録商標）である場合は、ファイル単位のコピーであればcpコマンド、またデバイス指定のブロック単位でのコピーであればddコマンドなどが挙げられる。

【0101】

また、一般的にオペレーティングシステムの設定により、これらのコマンドを定期的に行うように指定することができる。これにより共有LUを定期的にバックアップすることができる。さらに、管理端末160からなどの指示によって、ディスク制御部140が持つディスクコピー機能を利用して、CHNのオペレーティングシステムやコントローラマイクロプログラムとは無関係に、共有LU全体を共有LUバックアップにコピーすることも可能である。また、指示を契機にコピーを行うのではなく、CHN1が初めから共有LU311と共有LUバックアップ321に同時に共有データを書き込むことにより、共有LU311または共有LUバックアップ321が使用不能になった場合に残りの情報を使用してCHN5にフェイルオーバーをさせることも可能である。

【0102】

(バックアップデータの使用)

バックアップされたデータを使用する手順を以下に示す。

【0103】

共有LU311に障害が発生し、フェイルオーバなどの動作が出来なくなった場合、記憶装置システムに装備されている通報機能により保守員やシステム管理者に通報が行われる。保守員または管理者は、管理端末などより、共有LUバックアップ321上で、存在するファイルシステムをマウントする。その後、通常の処理をする。フェイルオーバが必要な場合は、共有LUバックアップ321上の情報を使用する。ついで保守員は障害が発生したドライブを交換するなどし、あたらしい共有LU311が準備できたら、再度ドライブを初期化し、共有LUバックアップ321から共有LU311に対し、バックアップを作成するのと同じ手段で逆方向にコピーする。

【0104】

(バックアップデータの格納先)

上記実施の形態では、バックアップは同一記憶装置内の別LUに作成したが、オペレーティングシステムからアクセスできる外部テープドライブなどに、NDMPプロトコルなどを使用してバックアップしてもよい。また、CHF1を経由して、SAN上のバックアップデバイスに対してバックアップしてもよい。さらに記憶装置のリモートコピー機能を利用して、別の記憶装置内へコピーしてもよい。

【0105】

次に、これらの処理手順を図に従って説明する。

【0106】

図17は、初めてNASをシステムに導入する際に、共有LUを初期化し、オペレーティングシステムをシステム用LUへインストールし、パーティションを作成するまでの手順を示している。図18は、CHNがフェイルオーバするときの手順を示している。図19は共有LUをバックアップし、共有LUが使用不能になった場合に共有LUバックアップの情報を用いてCHNがフェイルオーバするときの手順を示している。


【0107】

まず図17について説明する。図17の順番1から8までは、共有LUおよび共有LUバックアップをシステムとしてCHN1およびCHN5から認識できるようにするためのパス定義の手順である。次いで、図17の順番9から24までは、オペレーティングシステムをシステム用LUにインストールしながら、インストールソフトウェアによって共有LUおよび共有LUバックアップを初期化する手順を示している。

【0108】

システム管理者または保守員は、管理端末160から論理レベルでの共有LUの初期化を指示する(図17-順番1)。これにより共有LUは、ディスクアレイとして論理的に初期化される(図17-順番2)。この状態では共有LUは、パスが定義されればI/Oプロセッサ119からは読み書きが可能な状態になっているが、オペレーティングシステムからの認識はまだできる状態にはなっていない。ついでシステム管理者または保守員は、管理端末160から論理レベルでの共有LUバックアップの初期化を指示する(図17-順番3)。これにより共有LUバックアップは、ディスクアレイとして論理的に初期化される(図17-順番4)。その後システム管理者または保守員は、管理端末160から共有LUへのパス定義指示を行う(図17-順番5)。これにより、CHN1およびCHN5と共有LUとが関連付けられ、CHN1およびCHN5に属するI/Oプロセッサ119は、共有LUに対してアクセスすることが可能となる(図17-順番6)。さらにシステム管理者または保守員は、管理端末160から共有LUバックアップへのパス定義指示を行う(図17-順番7)。これにより、CHN1およびCHN5に属するI/Oプロセッサ119は、共有LUバックアップに対してアクセスすることが可能となる(図17-順番8)。こののちシステム管理者は、オペレーティングシステムのインストール指示を行う。

【0109】



まずシステム管理者または保守員は、オペレーティングシステムをCHN1にインストールするよう、管理端末160から指示を発行する(図17-順番9)。これによりCHN1のオペレーティングシステムのインストールが開始される(図17-順番10)。インストールソフトウェアは、CPU112上にロードされた後動作を開始し、他にこれまでCHNが存在せず、当該インストール動作がシステムで初めてのものであることを検出すると、共有LUをオペレーティングシステムで使用可能となるようにオペレーティングシステムレベルでの初期化を行う(図17-順番11)。この初期化指示は実際にはI/Oプロセッサ119を通じて行われる。またこの際、クラスタ毎に共有LUの所定の部分を使用することを予めソフトウェア的に決定してある場合は、インストールソフトウェアは、各クラスタが使用すべき共有LUの所定領域を割り当てるために、共有LUの領域を分割する。この処理をパーティション作成という(図17-順番11)。これにより共有LUは、I/Oプロセッサからのみならず、オペレーティングシステムからのアクセスが可能のように初期化され、かつそれぞれのクラスタに属するオペレーティングシステムがそれぞれ独自の領域にアクセスできるようにパーティションに分割される(図17-順番12)。同様に、インストールソフトウェアは、共有LUバックアップについてもオペレーティングシステムレベルでの初期化指示、パーティション分割指示を行い(図17-順番13)、共有LUバックアップは各クラスタに属するオペレーティングシステムからアクセス可能のように初期化され、パーティションに分割される(図17-順番14)。

【0110】

インストールソフトウェアは、引き続き共有LUの所定領域に、ファイルシステムを作成する(図17-順番15)。これはCHN1およびCHN5から共用されるものであるために、CHN1側で一旦作成されると、CHN5からは改めて作成する必要はない。この手順により共有LU上には、CHN1およびCHN5から、オペレーティングシステムがファイルシステムとしてアクセス可能な情報が作成される(図17-順番16)。同様に、インストールソフトウェアは、共有LUバックアップ上にファイルシステムを作成し(図17-順番17)、共有LUバックアップ上には、CHN1およびCHN5からオペレーティングシステムがファイルシステムとしてアクセス可能な情報が作成される(図17-順番18)。

【0111】

その後インストールソフトウェアは、CHN1のオペレーティングシステムを格納するLU領域に対しオペレーティングシステムのインストールを行い、それが完了すると管理端末に対してCHN1へのオペレーティングシステムのインストールが完了したことを通知する(図17-順番19)。管理端末160ではこの完了通知を受領すると(図17-順番20)、終了したことを示すメッセージを端末画面に出力する。システム管理者または保守員は当該メッセージを確認したのち、こんどはCHN5に対して同様にオペレーティングシステムのインストール指示を行う(図17-順番21)。CHN5ではインストールソフトウェアが実行され、オペレーティングシステムのネットワークインストールが開始される(図17-順番22)。ただし、ここでは、既にCHN1がシステムにインストール済みであるため、共有LUや共有LUバックアップの初期化は行わない。インストールソフトウェアは、CHN5のオペレーティングシステムを格納するLU領域に対しオペレーティングシステムのインストールを行い、それが完了すると管理端末に対してCHN5へのオペレーティングシステムのインストールが完了したことを通知する(図17-順番23)。管理端末160ではこの完了通知を受領すると(図17-順番24)、終了したことを示すメッセージを端末画面に出力する。これによって、システムへのオペレーティングシステムのインストールと、共有LUおよび共有LUバックアップの初期化、パーティション作成が完了する。

【0112】

なお、本実施例では、最初にオペレーティングシステムをインストールするCHN1上で動作するインストールソフトウェアが、他のCHNの使用するパーティションについてもすべて初期化を行ったが、そうではなくCHN1からはCHN1の使用する領域のみの

初期化を行い、各CHN上でオペレーティングシステムをインストールするタイミングで、それぞれ関連する領域を初期化する方法としてもよい。またCHN毎の領域は、ある一つの共有LU内のパーティション毎に分ける形式にしたが、そうではなくCHN毎に共有LUを独自に割り当て、各共有LUに他のCHNからもパスを定義しアクセスが可能とすることで情報を共有する形式としてもよい。

【0113】

次に、図18について説明する。

【0114】

図18は、CHN1のオペレーティングシステムが続行不能になった場合に、その業務をCHN5が引き継ぐ手順について述べている。

【0115】

まず、本手順ではCHN5のオペレーティングシステムは、CHN1の障害発生（図18-順番10）までのいずれかの時点で動作していればよいが、本実施例では、説明を簡単にするため、CHN1での動作説明をする前の時点から動作しているものとして扱う（図18-順番1）。

【0116】

CHN1上で動作しているオペレーティングシステムは、CHN1から使用する共有LUのファイルシステムをマウントする（図18-順番2）。このファイルシステムは、例えば図17-順番15で作成したファイルシステムである。マウントが完了すると、オペレーティングシステムが当該ファイルシステムに対して、データの読み書きを行うことができるようになる（図18-順番3）。こののちオペレーティングシステムは通常のクライアントに対するファイルサービスなどの処理を開始する（図18-順番4）。

【0117】

通常処理の間、CHN1のオペレーティングシステムは共有LUに対して、もしもCHN1が動作続行不能に陥った場合に、CHN5がクライアントに対するファイルサービスなどを肩代わりして再開できるような、引継ぎ情報を共有LUに書き込む（図18-順番5）。この書き込みは本実施例ではファイルシステムによる書き込みである。ファイルシステムを使用せず、たとえばブロックアクセスによって書き込む場合は、他のCHNからの書き込みと競合した場合の排他処理などを行う必要がある。また、引継ぎ情報は、クライアントのIPアドレスなどの情報、システム管理者や一般ユーザを含むユーザ情報、オペレーティングシステム上で動作しているサービスの動作情報、デーモンの動作情報など、あるいはユーザLUや共有LU、ファイルシステムをCHN1、CHNのどちらが使用しているかに行った情報、ファイルシステムがどのLUを使っているかと言う情報、CHN1やCHN5がクライアントに対して提供しているIPアドレスの情報などが含まれる。この引継ぎ情報の書き込みは、定期的あるいは必要情報に変更があった時点で、オペレーティングシステム自身の判断で共有LUに書き込まれる（図18-順番7、図18-順番9）。あるいは、管理端末160からの指示などによって、ユーザが共有LUに対して引継ぎ情報を書き込ませるようにしてもよい。

【0118】

これとは別に、CHN5上で動作しているオペレーティングシステムは、CHN1が引き続き動作しているかどうかを定期的に監視している（図18-順番6、図18-順番8、図18-順番11）。

【0119】

ここでCHN1側に何らかの障害が発生し、クライアントへのサービスが中断した（図18-順番10）とする。このとき、CHN5上で動作しているオペレーティングシステムからの動作監視（図18-順番11）は、CHN1で障害が発生したことを検出する（図18-順番12）。すると、CHN5上で動作しているオペレーティングシステムは、共有LU上のファイルシステムを自らマウントする（図18-順番13）。マウントが完了する（図18-順番14）と、CHN5のオペレーティングシステムは、それまでCHN1のオペレーティングシステムが使用していた引継ぎ情報にアクセスすることが可能と

なる。CHN5のオペレーティングシステムは、この引継ぎ情報を用いて、クライアントに対して自らがあたかもCHN1であるかのごとく振るまい、サービスを再開することが可能となる（図18－順番15）。

【0120】

なお、図18の実施例ではCHN1が通常処理系、CHN5が待機系という役割でフェイルオーバーを行う例を挙げた。この例はいわゆるアクティブスタンバイと呼ばれる動作形式であるが、この他にも、CHN1とCHN5双方が通常処理系の、いわゆるアクティブアクティブの形態を取ることもできる。その場合は各CHN毎に個別のファイルシステムを作成し、CHNの障害時には互いに相手が使用しているファイルシステムをマウントすることとしてもよい。また共有LUへの情報格納はファイルシステム形式ではなく、論理アドレス形式にして、各CHN毎に予めデータ格納エリアを論理アドレス毎に割り当てておく方式としてもよい。アクティブアクティブ形式の場合は、図18で示したような引継ぎ情報の書き込みや、CHNの動作状態の監視を相互に行うこととなる。

【0121】

次に、図19について説明する。

【0122】

図19は共有LUをバックアップし、共有LUが使用不能になった場合に共有LUバックアップの情報を用いてCHNがフェイルオーバーするときの手順を示している。

【0123】

図18と同様、CHN1で障害が発生した場合に処理を引き継ぐ役割のCHN5については、当初から通常処理を開始しているものとする（図19－順番1）。また、図18と同様、本図の例はアクティブスタンバイの形式の例であるが、CHN1とCHN5の役割を入れ替え、双方が行うことにすればアクティブアクティブの形態として運用することも可能である。

【0124】

CHN1上で動作しているオペレーティングシステムは、CHN1から使用する共有LUのファイルシステムをマウントする（図19－順番2）。このファイルシステムは、例えば図17－順番15で作成したファイルシステムである。マウントが完了すると、オペレーティングシステムが当該ファイルシステムに対して、データの読み書きを行うことができるようになる（図19－順番3）。こののちオペレーティングシステムは通常のクライアントに対するファイルサービスなどの処理を開始する（図19－順番4）。

【0125】

通常処理の間、CHN1のオペレーティングシステムは共有LUに対して、もしもCHN1が動作続行不能に陥った場合に、CHN5がクライアントに対するファイルサービスなどを肩代わりして再開できるような、引継ぎ情報を共有LUに書き込む（図19－順番5）。この書き込みは本実施例ではファイルシステムによる書き込みである。ファイルシステムを使用せず、たとえばブロックアクセスによって書き込む場合は、他のCHNからの書き込みと競合した場合の排他処理などを行う必要がある。また、引継ぎ情報は、クライアントのIPアドレスなどの情報、システム管理者や一般ユーザを含むユーザ情報、オペレーティングシステム上で動作しているサービスの動作情報、デーモンの動作情報など、あるいはユーザLUや共有LU、ファイルシステムをCHN1、CHNのどちらが使用しているかを行った情報、ファイルシステムがどのLUを使っているかと言う情報、CHN1やCHN5がクライアントに対して提供しているIPアドレスの情報などが含まれる。この引継ぎ情報の書き込みは、定期的あるいは必要情報に変更があった時点で、オペレーティングシステム自身の判断で共有LUに書き込まれる（図19－順番7）。あるいは、管理端末160からの指示などによって、ユーザが共有LUに対して引継ぎ情報を書き込ませるようにしてもよい。

【0126】

これとは別に、CHN5上で動作しているオペレーティングシステムは、CHN1が引き続き動作しているかどうかを定期的に監視している（図19－順番6、図19－順番1

6)。

【0127】

CHN1のオペレーティングシステムからはいつでも、OSのコピーコマンドにより、CHN1およびCHN5で使用する共有LUの領域をバックアップすることができる(図19-順番8)。これは例えば、LAN400上の情報処理装置200上で動作するUNIXオペレーティングシステム714や、Windowsオペレーティングシステム715から、CPU112上で動作しているオペレーティングシステム702にログインし、オペレーティングシステム702が汎用に提供しているコピーコマンドを使用してもよい。オペレーティングシステム702がUNIXである場合には、たとえばこのコマンドはcpというコマンドとなる。ファイルシステムを使用せず、LUデバイスを指定して直接データをブロック単位でコピーする場合には、このコマンドは例えばddというコマンドとなる。また、LAN400上の情報処理装置200上で動作するUNIXオペレーティングシステム714や、Windowsオペレーティングシステム715から、CPU112上で動作しているNASマネージャ等のアプリケーション706にログインし、アプリケーションの機能を用いてコピーしてもよい。あるいは管理端末160からこれらオペレーティングシステムやアプリケーションにログインしてコピー機能を実行してもよいし、管理端末160からディスク制御部140に指示を発行することにより、ディスク制御部が独自にコピーをおこなってもよい。更に、I/Oプロセッサ119上のコントローラマイクロプログラム、CPU112上のNASマネージャ等のアプリケーション、CPU112上のオペレーティングシステム、ディスク制御部140上のRAID制御部740やCPU142がシステム状態を監視し、たとえばシステムに対するデータ転送負荷の割合が一定値より低く、コピーを実行したとしてもクライアントに対するサービスに性能低下などの著しい悪影響を及ぼさないと判断した時点で、自動的にこれらのコマンドを起動および実行させ、バックアップを行うようにしてもよい。

【0128】

本実施例では、オペレーティングシステムが汎用に提供しているcpコマンドを情報処理端末200からログインして使うものとする。これにより、共有LU上のCHN1およびCHN5用の領域は共有LUバックアップの対応する領域にコピーされる(図19-順番9、図19-順番10)。

【0129】

また、バックアップを同一ディスクアレイ装置内の共有LUバックアップ321乃至324に対してではなく、外部バックアップデバイス900や、外部記憶装置システム610に対して作成してもよい。ここでは、NDMP(Network Data Management Protocol)を用いて、SAN経由でテープデバイス900に対して、CHN1およびCHN5用の共有LU領域をバックアップする方法を示す(図19-順番11)。これによりデータは外部バックアップデバイスにバックアップされる(図19-順番12、図19-順番13)。なお、外部記憶装置システム610に対して共有LUをバックアップした場合、その情報を用いて外部記憶装置システム610においてクライアントに対するサービスを引き継ぐことも考えられる。この場合はクライアントに対するサービスのみならず、クライアントがアクセスしていたユーザデータそのものもリモートコピー機能などを用いて外部記憶装置システム610に同期しておく必要がある。

【0130】

ここで、共有LUが障害などにより使用不能になったとする(図19-順番14)。この時点では、特にクライアントに対するファイルサービスに影響は発生しないが、ここで更にCHN1が何らかの障害が発生し(図19-順番15)、フェイルオーバーが必要になった場合に、図18-順番13のようなCHN5からの引継ぎ情報の取得が不可能となる。実際には、CHN5上のオペレーティングシステムはCHN1の障害発生を検出すると(図19-順番17)、共有LUのファイルシステムのマウント指示を行う(図19-順番18)が、このマウント操作は共有LUが使用不能であるため失敗する(図19-順番19)。この共有LUのマウント失敗を検出すると、CHN5のオペレーティングシステ

ムは、共有LUのバックアップにあるファイルシステムをマウントするように指示する（図19-順番20）。これにより共有LUのバックアップにあるファイルシステムに対して、CHN5のオペレーティングシステムが読み書きを行うことができるようになる（図19-順番21）。その後、CHN5のオペレーティングシステムは、共有LUのバックアップの引継ぎ情報を元に、クライアントに対してCHN1のファイルサービスなどの業務を再開することが可能となる（図19-順番22）。

【0131】

この後、例えば物理的なデバイスの交換などによって共有LUが再び使用可能となった場合は、デバイス交換の契機などによって共有LUを再び初期化し、図19-順番8などの手順を共有LUのバックアップから共有LUに対して実行することにより、再度共有LUに引継ぎ情報を書き戻すこともできる。

【0132】

以上本実施の形態について説明したが、上記実施の形態は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得るとともに、本発明にはその等価物等も含まれる。

【図面の簡単な説明】

【0133】

【図1】 本実施の形態に係る記憶装置システムの全体構成を示すブロック図である。

【図2】 本実施の形態に係る管理端末の構成を示すブロック図である。

【図3】 本実施の形態に係る物理ディスク管理テーブルを示す図である。

【図4】 本実施の形態に係るLU管理テーブルを示す図である。

【図5】 本実施の形態に係る記憶装置システムの外観構成を示す図である。

【図6】 本実施の形態に係る記憶デバイス制御装置の外観構成を示す図である。

【図7】 本実施の形態に係るCHNのハードウェア構成を示す図である。

【図8】 本実施の形態に係るメモリに記憶されるデータの内容を説明するための図である。

【図9】 本実施の形態に係るメタデータを示す図である。

【図10】 本実施の形態に係るロックデータを示す図である。

【図11】 本実施の形態に係るCHN上のCPUとI/Oプロセッサとの通信経路を示す図である。

【図12】 本実施の形態に係るCHN上の内部LANを介したハードウェア構成を示す図である。

【図13】 本実施の形態に係るディスク制御部を示す図である。

【図14】 本実施の形態に係る記憶装置システムのソフトウェア構成図である。

【図15】 本実施の形態に係るオペレーティングシステム用LU、共有LUの論理構成を示す図である。

【図16】 本実施の形態に係る共有LUをパーティション毎に分けてバックアップする場合の論理構成を示す図である。

【図17】 本実施の形態に係る共有LUを初期化しパーティションごとに分ける手順を示す図である。

【図18】 本実施の形態に係る共有LUに格納した情報を用いてCHNがフェイルオーバーする手順を示す図である。

【図19】 本実施の形態に係る共有LUをバックアップする手順、および共有LUが使用不能になった場合に共有LUバックアップの情報を用いてCHNがフェイルオーバーする手順を示す図である。

【符号の説明】

【0134】

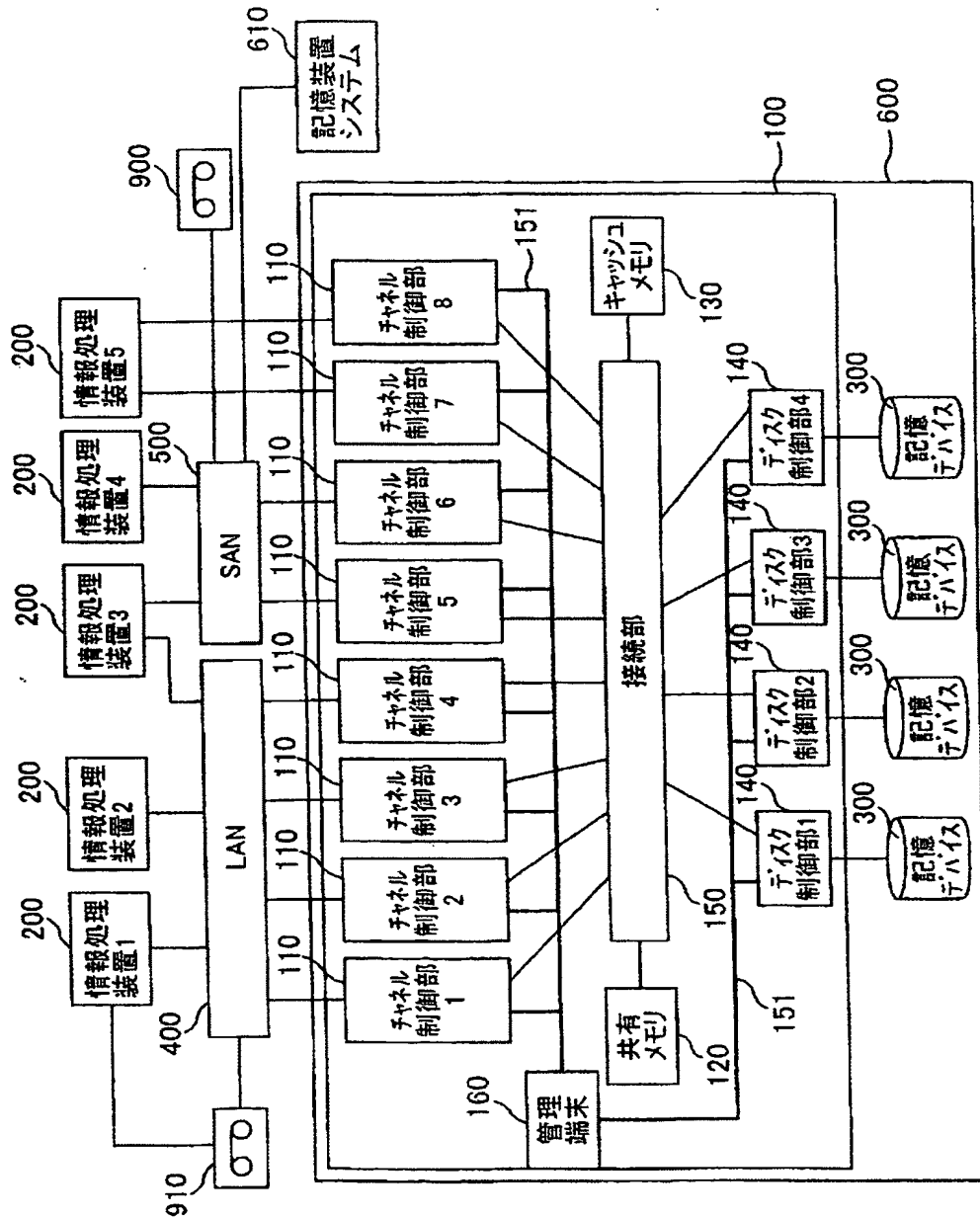
- 100 記憶デバイス制御装置
- 110 チャネル制御部
- 111 ネットワークインタフェース部

1 1 2 C P U
1 1 3 メモリ
1 1 4 入出力制御部
1 1 5 N V R A M
1 1 6 ボード接続用コネクタ
1 1 7 通信コネクタ
1 1 8 回路基板
1 1 9 I / O プロセッサ
1 2 0 共有メモリ
1 3 0 キャッシュメモリ
1 4 0 ディスク制御部
1 5 0 接続部
1 5 1 内部 L A N
1 6 0 管理端末
6 0 0 記憶装置システム
8 0 1 B I O S
8 0 2 通信メモリ
8 0 3 ハードウェアレジスタ群
8 0 4 N V R A M

【書類名】 図面

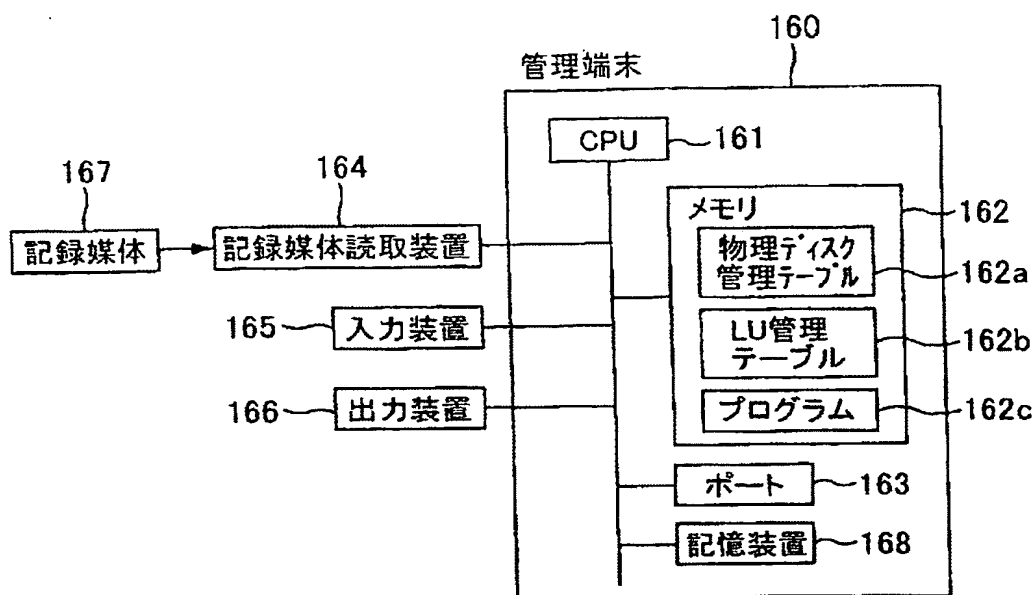
【図1】

【図1】



【図 2】

【図 2】



【図 3】

【図 3】

162a

物理ディスク管理テーブル

ディスク番号	容量	RAID	使用状況
#001	100GB	5	使用中
#002	100GB	5	使用中
#003	100GB	5	使用中
#004	100GB	5	使用中
#005	100GB	5	使用中
#006	50GB	—	未使用
⋮	⋮	⋮	⋮

【図 4】

【図 4】

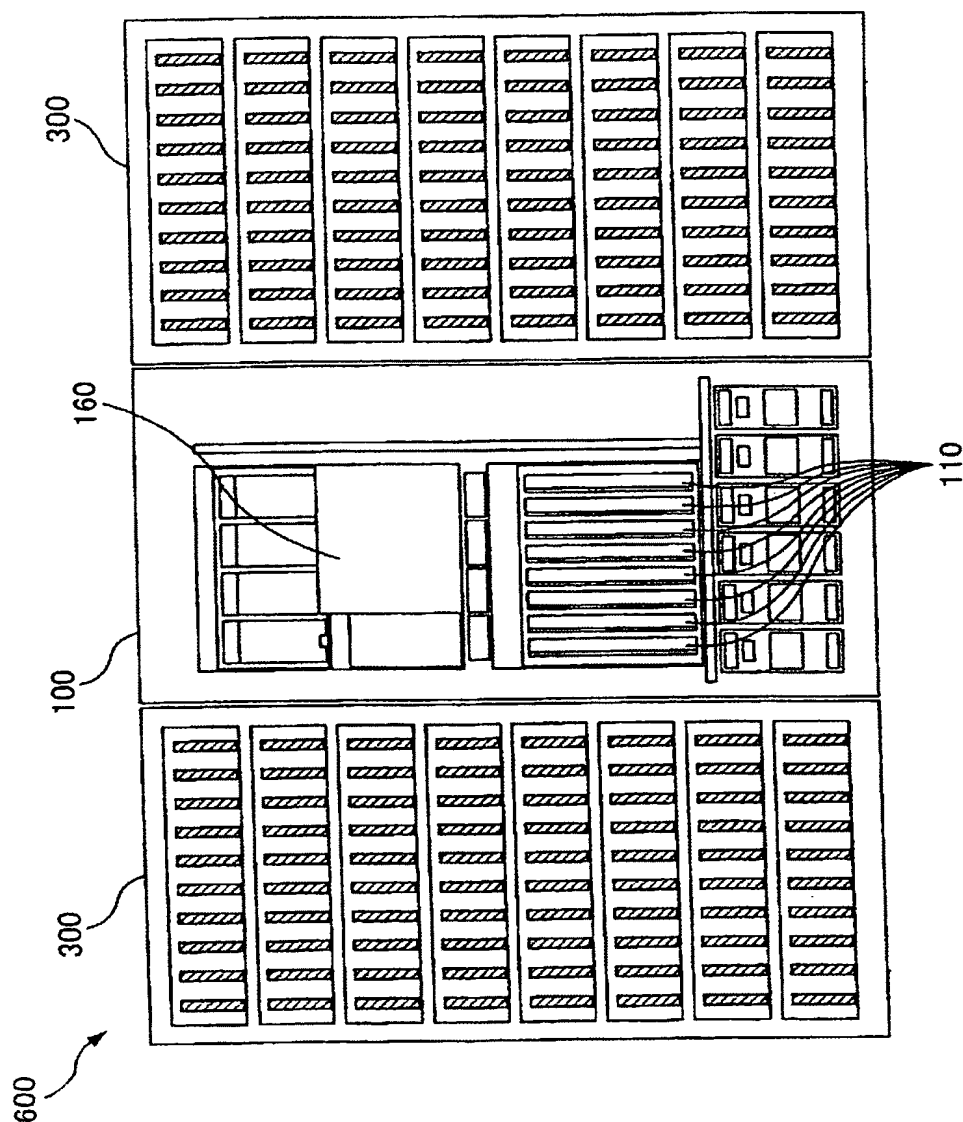
162b

LU管理テーブル

LU番号	物理ディスク	容量	RAID
#1	#001,#002,#003,#004,#005	100GB	5
#2	#001,#002,#003,#004,#005	300GB	5
#3	#006,#007,	200GB	1
⋮	⋮	⋮	⋮

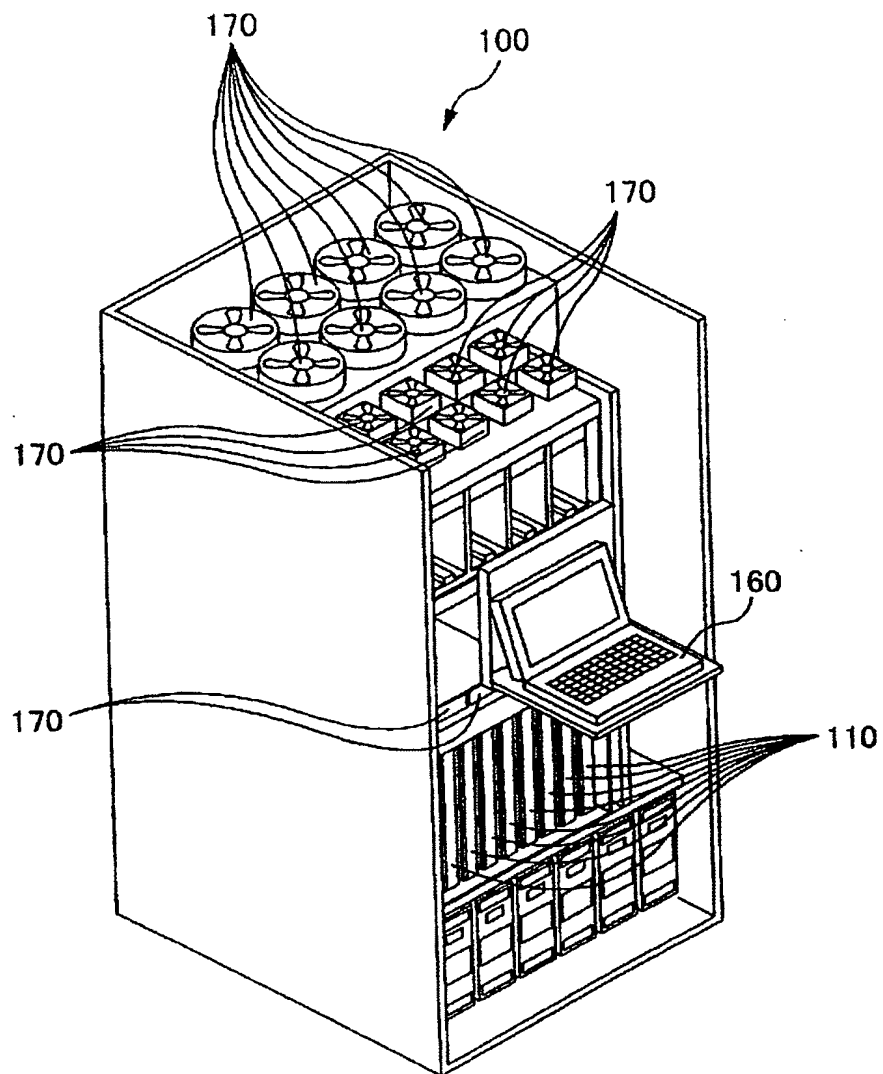
【図 5】

【図 5】



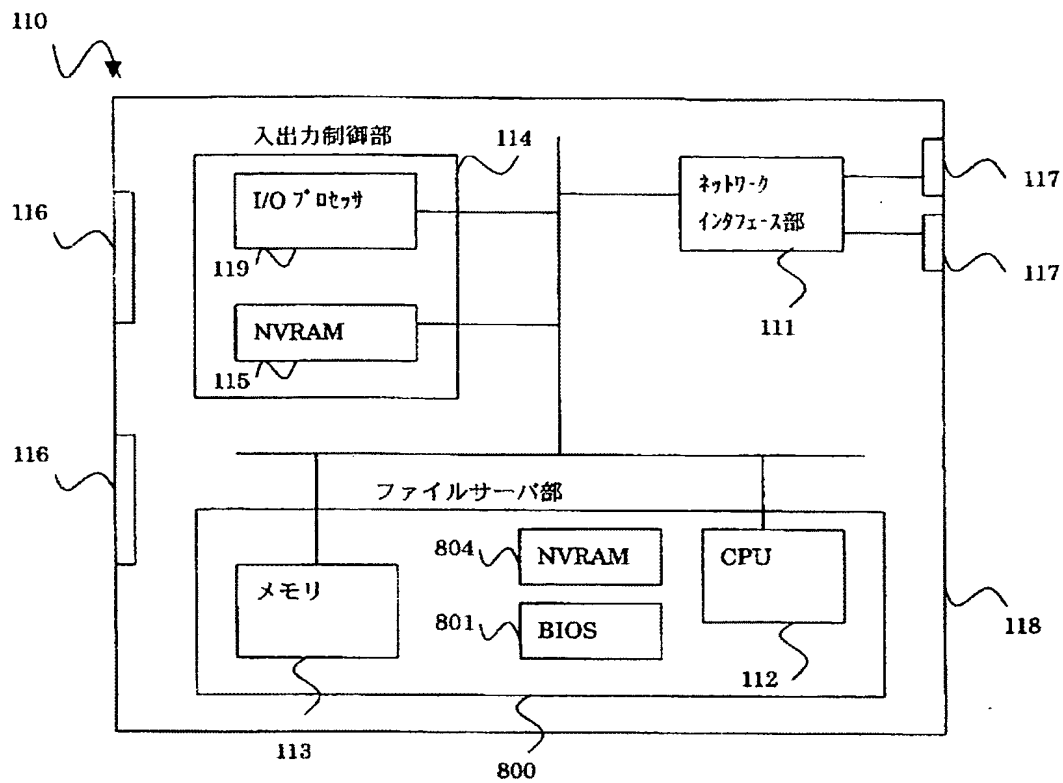
【図 6】

【図 6】



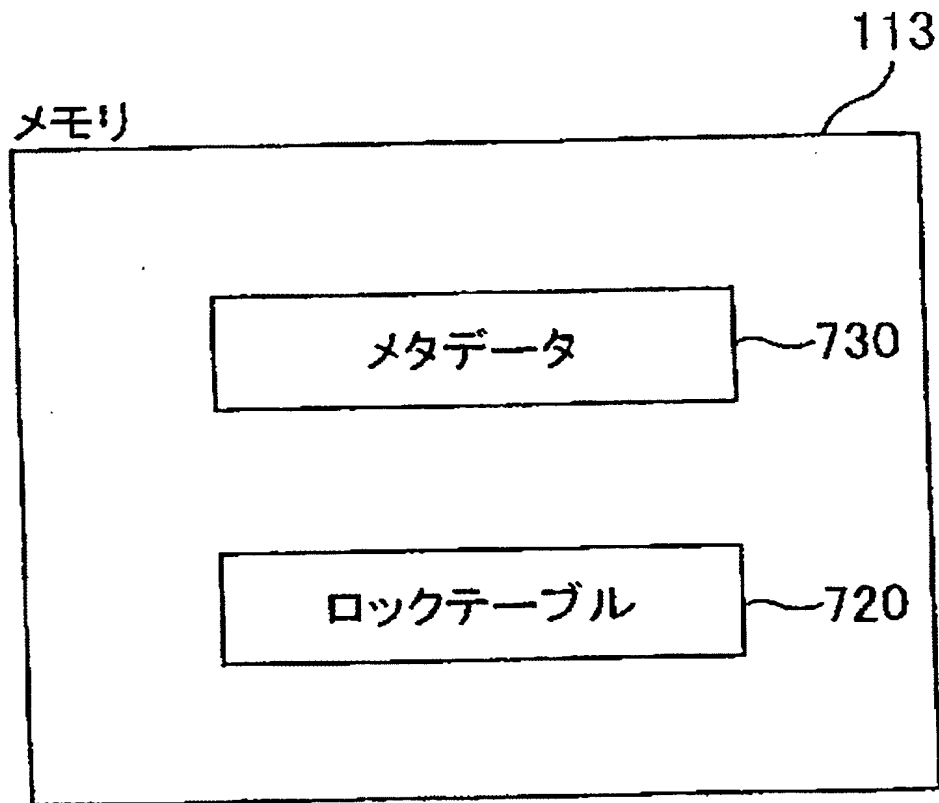
【図 7】

【図 7】



【図 8】

【図 8】



【図 9】

【図 9】

730

メタデータ

ファイル名	先頭アドレス	容量	所有者	更新時刻
A	7BSA	200MB	X	0:00
B	05BF	50MB	X	7:57
C	1F30	100MB	Y	9:15
D	470B	100MB	Z	15:20
⋮	⋮	⋮	⋮	⋮

【図 10】

【図 10】

721

ファイルロックテーブル

ファイル名	ロック状態
A	ロック中
B	—
C	—
D	ロック中
⋮	⋮

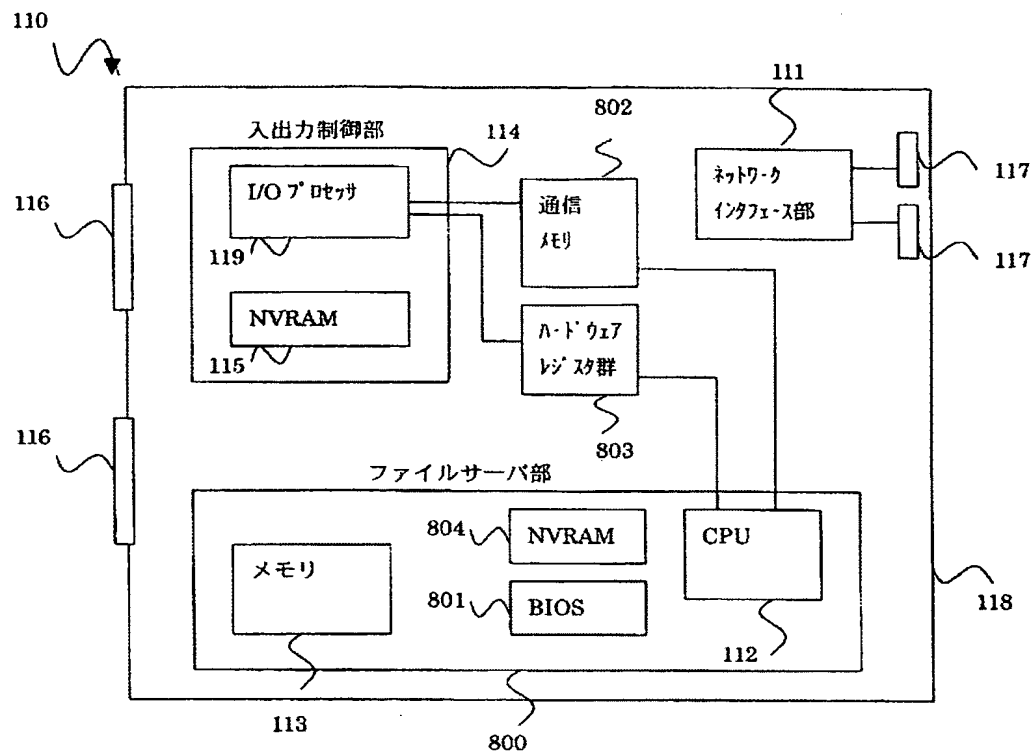
722

LUロックテーブル

LU	ロック状態
共有	—
1	ロック中
2	—
⋮	⋮

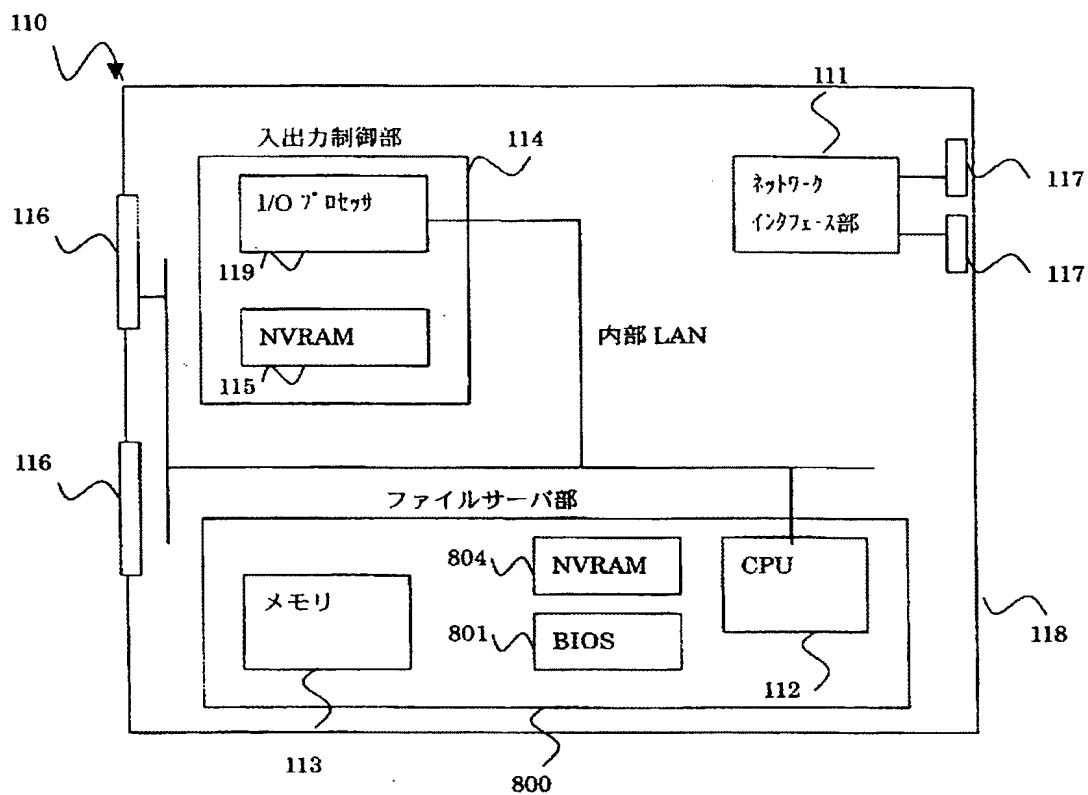
【図 11】

【図 11】



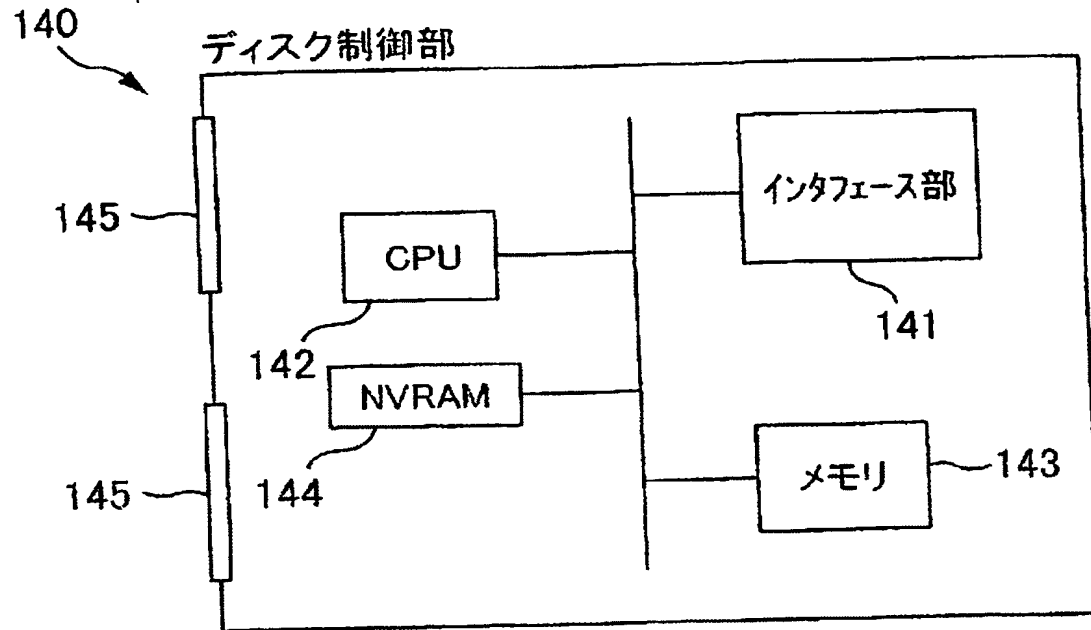
【図 12】

【図 12】



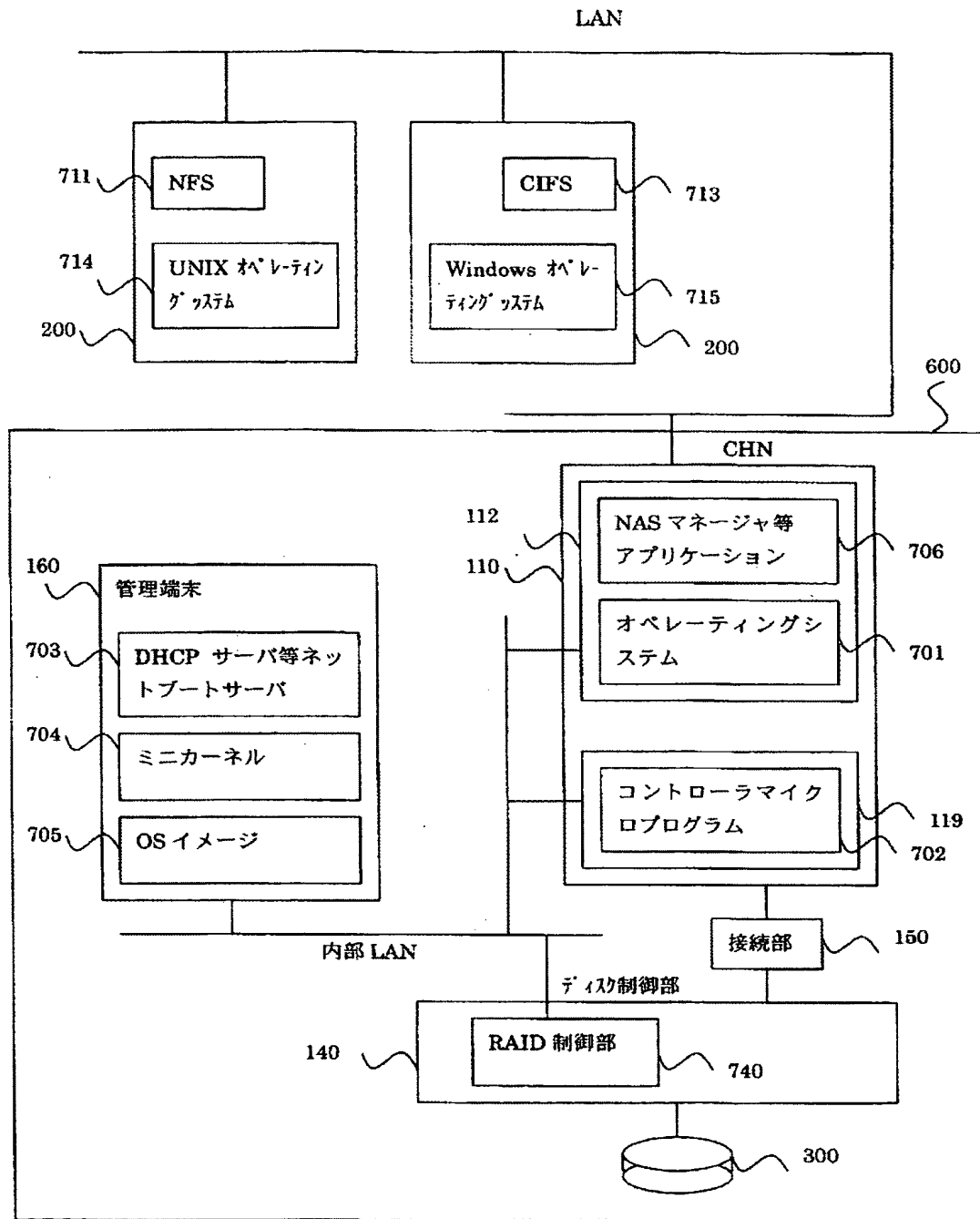
【図 13】

【図 13】

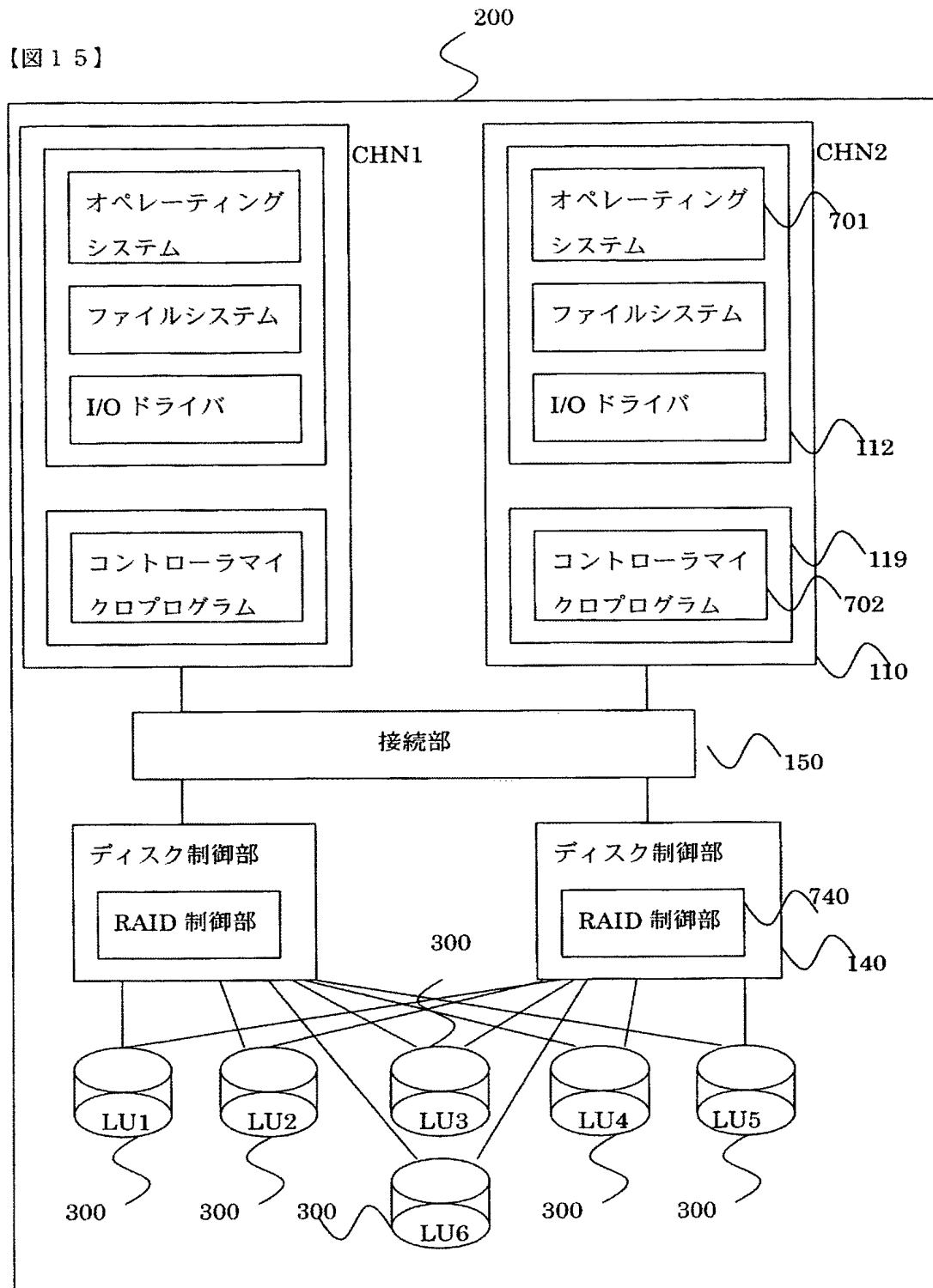


【図 14】

【図 14】

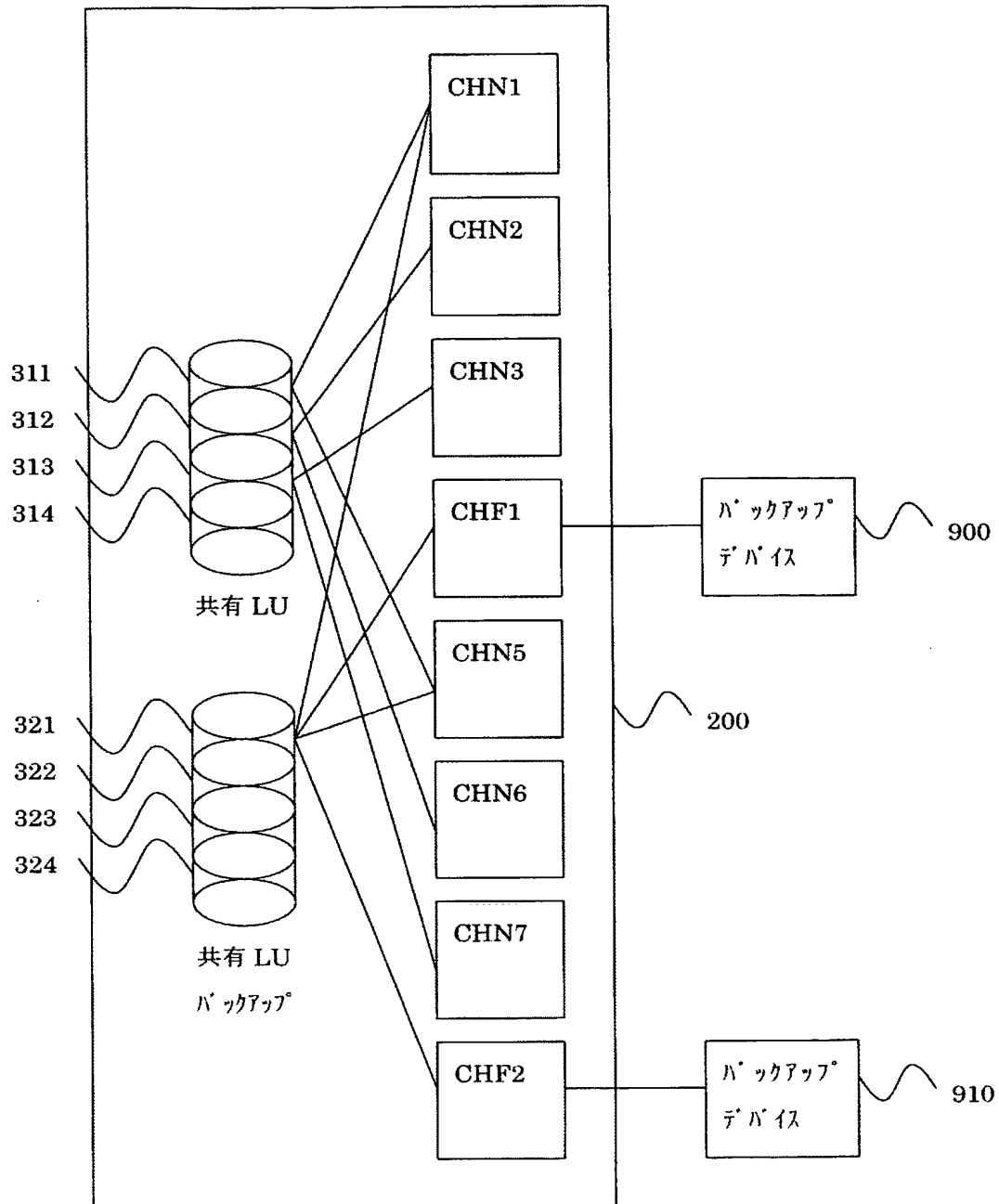


【図 15】



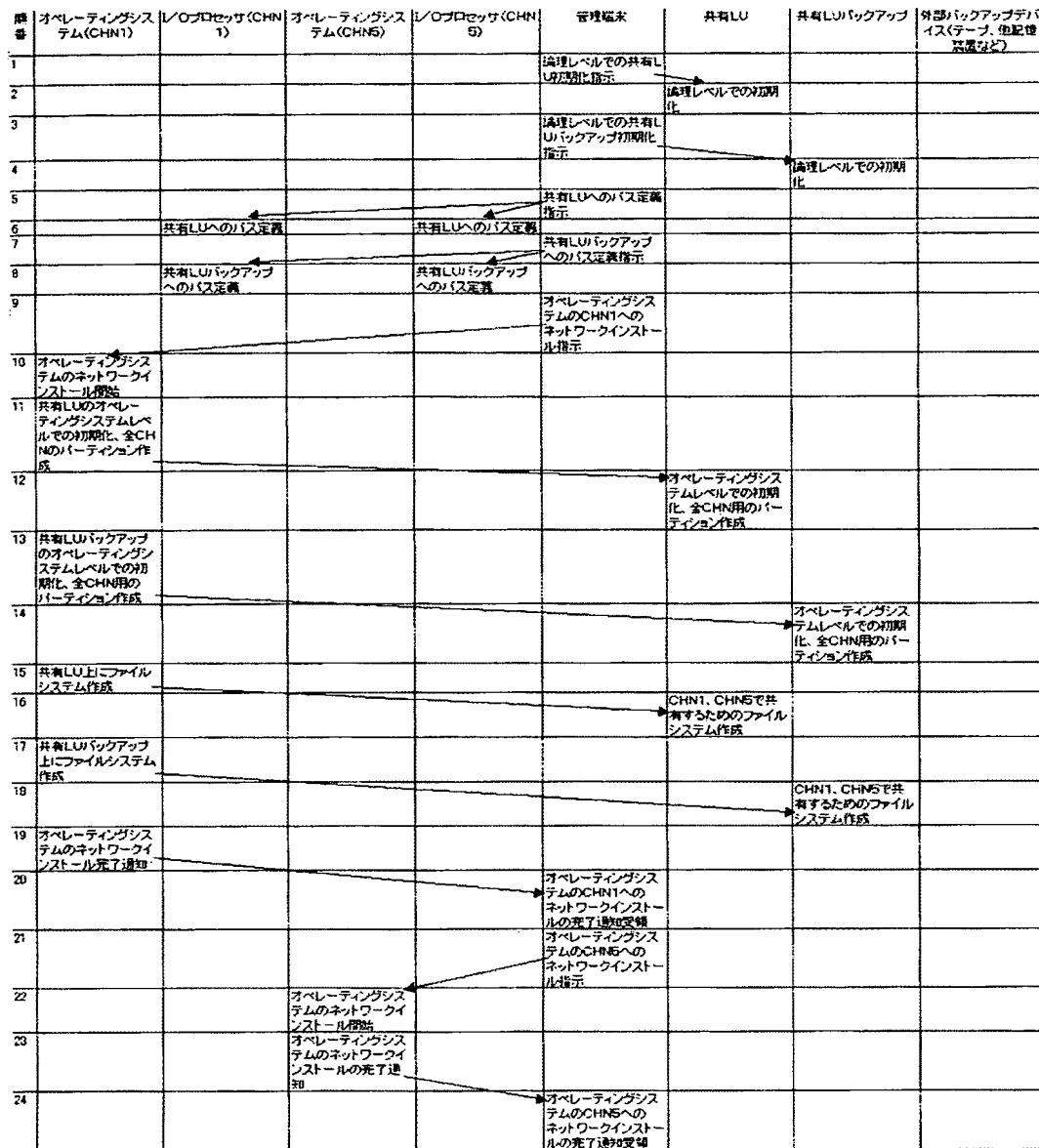
【図 16】

【図 16】



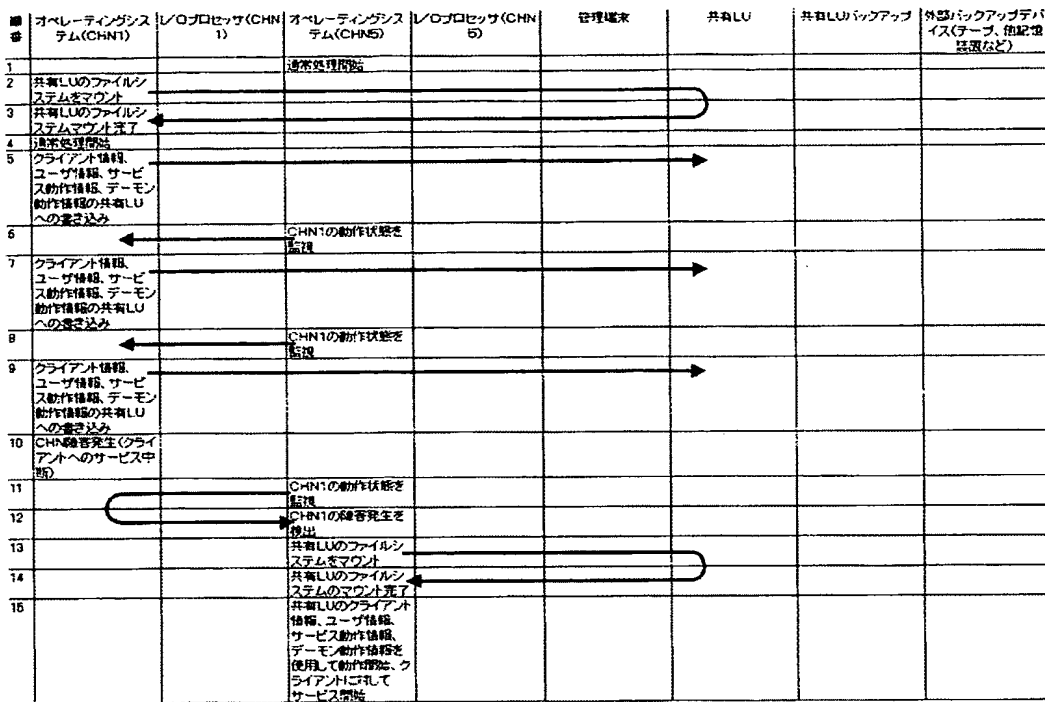
【図 17】

【図 17】



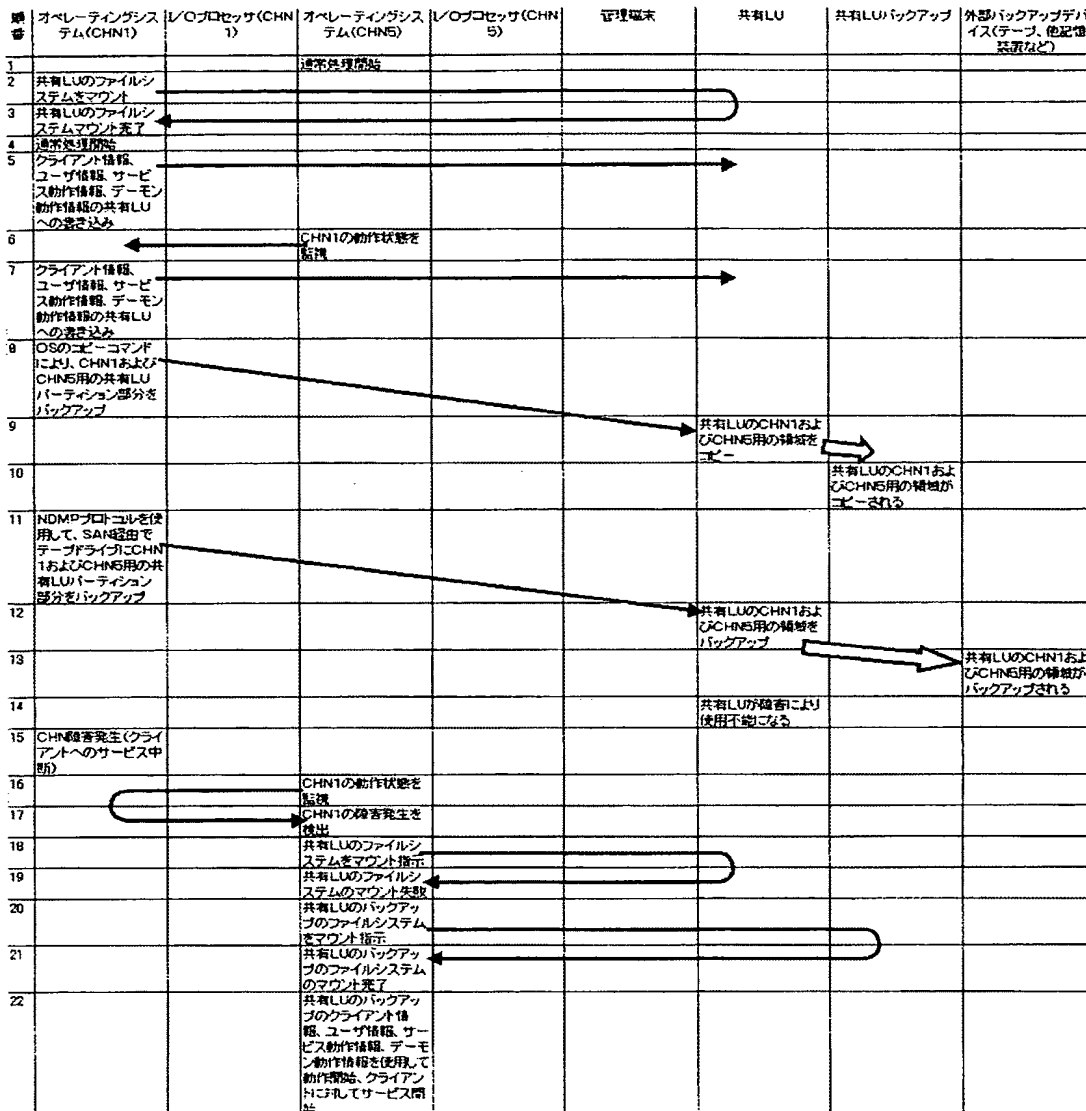
【図 18】

【図 18】



【図 19】

【図 19】



【書類名】 要約書**【要約】****【課題】**

複数の異種ネットワークに接続可能な記憶装置システムを提供するとともに、かかる記憶装置システムを発明するにあたり必要とされる記憶デバイス制御装置、及びデバイス制御装置の起動を制御する方法を提供する。

【解決手段】

ディスクアレイ装置は、複数の記憶デバイスと、記憶デバイス制御部と、前記記憶デバイス制御部に接続される接続部と、複数の第一のチャンネル制御部と、共有メモリと、キャッシュメモリと、を有する。

第一のチャンネル制御部は、自ディスクアレイ装置の外部のローカルエリアネットワークを介して受けたファイルレベルのデータをブロックレベルのデータに変換して、複数の記憶デバイスへの格納を要求する第一のプロセッサと、第一のプロセッサからの要求に応じて接続部及び記憶デバイス制御部を介して複数の記憶デバイスへ前記ブロックレベルのデータを転送する第二のプロセッサとを有し、前記接続部及び前記ローカルエリアネットワークに接続される。

前記複数の第一のチャンネル制御部内の前記第二のプロセッサは、ブロックレベルのデータが格納される複数の記憶領域と、複数の第一のプロセッサによって相互にやり取りされるプロセッサ間の処理状況に関する情報が格納されるプロセッサ情報格納領域と、を前記複数の記憶デバイスの記憶領域を用いて作成する。

【選択図】 図 1 6

認定・付加情報

特許出願の番号	特願 2 0 0 3 - 3 9 4 9 2 2
受付番号	5 0 3 0 1 9 4 1 3 2 9
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 5 年 1 1 月 2 7 日

< 認定情報・付加情報 >

【提出日】	平成 15 年 11 月 26 日
-------	-------------------

特願 2 0 0 3 - 3 9 4 9 2 2

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 5 1 0 8]

1. 変更年月日	1 9 9 0 年 8 月 3 1 日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台 4 丁目 6 番地
氏 名	株式会社日立製作所